

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds
(without alignments)
800.703 Million cell updates/sec

Title: US-10-681-773-1
Perfect score: 25
Sequence: 1 aaaaaaaaaatcgcaacaatc 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/1_COMB.seq:*\n2: /cgn2_6/prodata/1/ina/5_COMB.seq:*\n3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*\n4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*\n5: /cgn2_6/prodata/1/ina/H_COMB.seq:*\n6: /cgn2_6/prodata/1/ina/PCrUS_COMB.seq:*\n7: /cgn2_6/prodata/1/ina/PP_COMB.seq:*\n8: /cgn2_6/prodata/1/ina/RE_COMB.seq:*\n9: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23.4	93.6	40168	3	US-09-949-016-13225
C 2	23.4	93.6	46841	3	US-09-949-016-13466
C 3	23.4	93.6	60141	3	US-09-949-016-15874
C 4	23.4	93.6	360470	3	US-09-949-016-13173
5	21.8	87.2	91772	3	US-09-949-016-15332
6	21.8	87.2	91772	3	US-09-949-016-15332
7	21.8	87.2	91772	3	US-09-949-016-15332
8	20.8	83.2	488	3	US-09-270-767-24163
9	20.8	83.2	488	3	US-09-270-767-24163
C 10	20.8	83.2	601	3	US-09-949-016-105783
C 11	20.8	83.2	1834	3	US-09-949-016-14688
C 12	20.8	83.2	86857	3	US-09-949-016-14688
C 13	20.4	81.6	601	3	US-09-949-016-105775
C 14	20.4	81.6	52655	3	US-09-949-016-13495
C 15	20.4	81.6	86857	3	US-09-949-016-14688
16	20.4	81.6	92387	3	US-09-949-016-14563
17	20.4	81.6	151295	3	US-09-949-016-14563
18	20.4	81.6	151295	3	US-09-949-016-14563
19	20.4	81.6	151295	3	US-09-949-016-14563
20	20.4	81.6	151295	3	US-09-949-016-14563
21	20.4	81.6	151295	3	US-09-949-016-14563
22	20.4	81.6	151295	3	US-09-949-016-14563
C 23	20.4	81.6	393753	3	US-09-949-016-14573
C 24	20.4	81.6	393753	3	US-09-949-016-14573

25	20.4	81.6	818128	3	US-09-949-016-14546	Sequence 14546, A
26	20.4	81.6	818128	3	US-09-949-016-14547	Sequence 14547, A
27	20.4	81.6	818128	3	US-09-949-016-14548	Sequence 14548, A
28	20.4	81.6	818128	3	US-09-949-016-14549	Sequence 14549, A
29	20.4	81.6	818128	3	US-09-949-016-14550	Sequence 14550, A
30	20.4	81.6	818128	3	US-09-949-016-14551	Sequence 14551, A
31	20.4	81.6	818128	3	US-09-949-016-14552	Sequence 14552, A
32	20.4	81.6	818128	3	US-09-949-016-14553	Sequence 14553, A
33	20.4	81.6	818128	3	US-09-949-016-14554	Sequence 14554, A
34	20.4	81.6	818128	3	US-09-949-016-14555	Sequence 14555, A
35	20.4	81.6	818128	3	US-09-949-016-14556	Sequence 14556, A
36	20.4	81.6	818128	3	US-09-949-016-14557	Sequence 14557, A
37	20.4	81.6	818128	3	US-09-949-016-14558	Sequence 14558, A
38	20.4	81.6	818128	3	US-09-949-016-14559	Sequence 14559, A
39	20.4	81.6	818128	3	US-09-949-016-14560	Sequence 14560, A
40	20.4	81.6	818128	3	US-09-949-016-14561	Sequence 14561, A
41	20.4	81.6	818128	3	US-09-949-016-14562	Sequence 14562, A
42	20.4	81.6	818128	3	US-09-949-016-14563	Sequence 14563, A
43	20.4	81.6	818128	3	US-09-949-016-14564	Sequence 14564, A
44	20.4	81.6	818128	3	US-09-949-016-14565	Sequence 14565, A
45	20.4	81.6	818128	3	US-09-949-016-14566	Sequence 14566, A
46	20.2	80.8	601	3	US-09-949-016-14567	Sequence 14567, A
C 47	20.2	80.8	601	3	US-09-949-016-14568	Sequence 14568, A
C 48	20.2	80.8	601	3	US-09-949-016-14569	Sequence 14569, A
49	20.2	80.8	601	3	US-09-949-016-14570	Sequence 14570, A
50	20.2	80.8	601	3	US-09-949-016-14571	Sequence 14571, A
51	20.2	80.8	601	3	US-09-949-016-14572	Sequence 14572, A
52	20.2	80.8	601	3	US-09-949-016-14573	Sequence 14573, A
53	20.2	80.8	601	3	US-09-949-016-14574	Sequence 14574, A
C 54	20.2	80.8	601	3	US-09-949-016-14575	Sequence 14575, A
55	20.2	80.8	601	3	US-09-949-016-14576	Sequence 14576, A
56	20.2	80.8	601	3	US-09-949-016-14577	Sequence 14577, A
57	20.2	80.8	601	3	US-09-949-016-14578	Sequence 14578, A
58	20.2	80.8	601	3	US-09-949-016-14579	Sequence 14579, A
59	20.2	80.8	601	3	US-09-949-016-14580	Sequence 14580, A
60	20.2	80.8	601	3	US-09-949-016-14581	Sequence 14581, A
61	20.2	80.8	601	3	US-09-949-016-14582	Sequence 14582, A
62	20.2	80.8	601	3	US-09-949-016-14583	Sequence 14583, A
63	20.2	80.8	601	3	US-09-949-016-14584	Sequence 14584, A
64	20.2	80.8	601	3	US-09-949-016-14585	Sequence 14585, A
C 65	20.2	80.8	601	3	US-09-949-016-14586	Sequence 14586, A
C 66	20.2	80.8	601	3	US-09-949-016-14587	Sequence 14587, A
C 67	20.2	80.8	601	3	US-09-949-016-14588	Sequence 14588, A
C 68	20.2	80.8	601	3	US-09-949-016-14589	Sequence 14589, A
C 69	20.2	80.8	601	3	US-09-949-016-14590	Sequence 14590, A
C 70	20.2	80.8	601	3	US-09-949-016-14591	Sequence 14591, A
C 71	20.2	80.8	601	3	US-09-949-016-14592	Sequence 14592, A
72	20.2	80.8	601	3	US-09-949-016-14593	Sequence 14593, A
73	20.2	80.8	601	3	US-09-949-016-14594	Sequence 14594, A
C 74	20.2	80.8	601	3	US-09-949-016-14595	Sequence 14595, A
C 75	20.2	80.8	601	3	US-09-949-016-14596	Sequence 14596, A
C 76	20.2	80.8	601	3	US-09-949-016-14597	Sequence 14597, A
C 77	20.2	80.8	601	3	US-09-949-016-14598	Sequence 14598, A
78	20.2	80.8	601	3	US-09-949-016-14599	Sequence 14599, A
79	20.2	80.8	601	3	US-09-949-016-14600	Sequence 14600, A
C 80	20.2	80.8	601	3	US-09-949-016-14601	Sequence 14601, A
C 81	20.2	80.8	601	3	US-09-949-016-14602	Sequence 14602, A
C 82	20.2	80.8	601	3	US-09-949-016-14603	Sequence 14603, A
C 83	20.2	80.8	601	3	US-09-949-016-14604	Sequence 14604, A
84	20.2	80.8	601	3	US-09-949-016-14605	Sequence 14605, A
85	20.2	80.8	601	3	US-09-949-016-14606	Sequence 14606, A
86	20.2	80.8	601	3	US-09-949-016-14607	Sequence 14607, A
87	20.2	80.8	601	3	US-09-949-016-14608	Sequence 14608, A
88	20.2	80.8	601	3	US-09-949-016-14609	Sequence 14609, A
89	20.2	80.8	601	3	US-09-949-016-14610	Sequence 14610, A
C 90	20.2	80.8	601	3	US-09-949-016-14611	Sequence 14611, A
91	20.2	80.8	601	3	US-09-949-016-14612	Sequence 14612, A
92	20.2	80.8	601	3	US-09-949-016-14613	Sequence 14613, A
93	20.2	80.8	601	3	US-09-949-016-14614	Sequence 14614, A
94	20.2	80.8	601	3	US-09-949-016-14615	Sequence 14615, A
C 95	19.8	79.2	520	3	US-09-949-016-14616	Sequence 14616, A
C 96	19.8	79.2	520	3	US-09-949-016-14617	Sequence 14617, A
97	19.8	79.2	520	3	US-09-949-016-14618	Sequence 14618, A

```
98 19.8 79.2 601 3 US-09-949-016-78415 Sequence 78415, A
C 99 19.8 79.2 601 3 US-09-949-016-104950 Sequence 104950, A
C 100 19.8 79.2 601 3 US-09-949-016-104951 Sequence 104951, A
C 101 19.8 79.2 601 3 US-09-949-016-104975 Sequence 104975, A
C 102 19.8 79.2 601 3 US-09-949-016-104976 Sequence 104976, A
C 103 19.8 79.2 601 3 US-09-949-016-10821 Sequence 10821, A
C 104 19.8 79.2 36093 3 US-09-949-016-14664 Sequence 14664, A
C 105 19.8 79.2 36093 3 US-09-949-016-14665 Sequence 14665, A
C 106 19.8 79.2 331814 3 US-09-949-016-12008 Sequence 12008, A
C 107 19.8 79.2 331814 3 US-09-949-016-17056 Sequence 17056, A
C 108 19.4 77.6 601 3 US-09-949-016-29513 Sequence 29513, A
C 109 19.4 77.6 601 3 US-09-949-016-53066 Sequence 53066, A
C 110 19.4 74527 3 US-09-949-016-12339 Sequence 12339, A
C 111 19.4 77.6 74528 3 US-09-949-016-13275 Sequence 13275, A
C 112 19.2 76.8 243 3 US-09-248-796A-13226 Sequence 13236, A
C 113 19.2 76.8 516 3 US-09-302-626B-79 Sequence 79, Appl
C 114 19.2 76.8 601 3 US-09-949-016-18123 Sequence 18123, A
C 115 19.2 76.8 601 3 US-09-949-016-18124 Sequence 18124, A
C 116 19.2 76.8 601 3 US-09-949-016-40296 Sequence 40296, A
C 117 19.2 76.8 601 3 US-09-949-016-61144 Sequence 61144, A
C 118 19.2 76.8 601 3 US-09-949-016-61145 Sequence 61145, A
C 119 19.2 76.8 601 3 US-09-949-016-71199 Sequence 71199, A
C 120 19.2 76.8 601 3 US-09-949-016-105610 Sequence 105610, A
C 121 19.2 76.8 601 3 US-09-949-016-105611 Sequence 105611, A
C 122 19.2 76.8 601 3 US-09-949-016-137303 Sequence 137303, A
C 123 19.2 76.8 601 3 US-09-949-016-147298 Sequence 147298, A
C 124 19.2 76.8 601 3 US-09-949-016-156873 Sequence 156873, A
C 125 19.2 76.8 601 3 US-09-949-002-5335 Sequence 5335, Ap
C 126 19.2 76.8 729 3 US-09-302-626B-81 Sequence 81, Appl
C 127 19.2 76.8 729 3 US-09-302-626B-83 Sequence 83, Appl
C 128 19.2 76.8 1367 3 US-09-270-767-3617 Sequence 3617, Ap
C 129 19.2 76.8 1367 3 US-09-270-767-18899 Sequence 18899, A
C 130 19.2 76.8 1455 3 US-09-248-796A-4726 Sequence 4726, Ap
C 131 19.2 76.8 6669 3 US-09-949-016-17534 Sequence 17534, A
C 132 19.2 76.8 11988 3 US-09-949-016-11977 Sequence 11977, A
C 133 19.2 76.8 11989 3 US-09-949-016-13676 Sequence 13676, A
C 134 19.2 76.8 13711 3 US-09-949-016-11798 Sequence 11798, A
C 135 19.2 76.8 15702 3 US-09-949-016-15518 Sequence 15518, A
C 136 19.2 76.8 24953 3 US-09-949-016-15743 Sequence 15743, A
C 137 19.2 76.8 29569 3 US-09-949-016-15875 Sequence 15875, A
C 138 19.2 76.8 30291 3 US-09-949-016-18875 Sequence 18875, A
C 139 19.2 76.8 30649 3 US-09-949-002-643 Sequence 643, App
C 140 19.2 76.8 30649 3 US-09-949-002-723 Sequence 723, App
C 141 19.2 76.8 43950 3 US-09-735-934A-3 Sequence 3, Appli
C 142 19.2 76.8 43950 3 US-10-060-332-3 Sequence 3, Appli
C 143 19.2 76.8 43950 3 US-10-339-657-3 Sequence 3, Appli
C 144 19.2 76.8 43950 3 US-10-885-879-3 Sequence 3, Appli
C 145 19.2 76.8 45684 3 US-09-949-016-16539 Sequence 16539, A
C 146 19.2 76.8 45684 3 US-09-949-016-11767 Sequence 11767, A
C 147 19.2 76.8 45639 3 US-09-949-016-13520 Sequence 13520, A
C 148 19.2 76.8 66175 3 US-09-949-016-12293 Sequence 12293, A
C 149 19.2 76.8 77618 3 US-09-949-016-11768 Sequence 11768, A
C 150 19.2 76.8 81335 3 US-09-949-002-709 Sequence 709, App
```

ALIGNMENTS

```
RESULT 1
US-09-949-016-13225/c
; Sequence 13225, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13225
; LENGTH: 40168
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(40168)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13225
```

```
Query Match 93.6%; Score 23.4; DB 3; Length 40168;
Best Local Similarity 96.0%; Pred. No. 25;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAATCT 25
Db 8976 AAAAAAAAAATCGCAACAATCT 8952
```

```
RESULT 2
US-09-949-016-13466/c
; Sequence 13466, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13466
; LENGTH: 46841
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13466
```

```
Query Match 93.6%; Score 23.4; DB 3; Length 46841;
Best Local Similarity 96.0%; Pred. No. 25;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAATCT 25
Db 20040 AAAAAAAAAATCGCAACAATCT 20016
```

```
RESULT 3
US-09-949-016-15874/c
; Sequence 15874, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```


;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 15874
;; LENGTH: 60141
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(60141)
;; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15874

Query Match 93.6%; Score 23.4; DB 3; Length 60141;
Best Local Similarity 96.0%; Pred. No. 25;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
DB 47220 AAAAAAAAAATCGCAACAATCT 47196

RESULT 4
US-09-949-016-13173/C
;; Sequence 13173, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 13173
;; LENGTH: 360470
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-13173

Query Match 93.6%; Score 23.4; DB 3; Length 360470;
Best Local Similarity 96.0%; Pred. No. 25;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
DB 301343 AAAAAAAAAATCGCAACAATCT 301319

RESULT 5
US-09-949-016-15332
;; Sequence 15332, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08

;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 15332
;; LENGTH: 71251
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-15332

Query Match 87.2%; Score 21.8; DB 3; Length 71251;
Best Local Similarity 92.0%; Pred. No. 83;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
DB 50265 AAAAAAAAAATCGCAACAATCT 50289

RESULT 6
US-09-949-016-15568
;; Sequence 15568, Application US/09949016
;; Patent No. 6812339
;; GENERAL INFORMATION:
;; APPLICANT: VENTER, J. Craig et al.
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; PRIOR FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 15568
;; LENGTH: 91772
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(91772)
;; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15568

Query Match 87.2%; Score 21.8; DB 3; Length 91772;
Best Local Similarity 92.0%; Pred. No. 83;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
DB 48082 AAAAAAAAAATCGCAACAATCT 48106

RESULT 7
US-09-573-080A-324
;; Sequence 324, Application US/09573080A
;; Patent No. 6828057
;; GENERAL INFORMATION:
;; APPLICANT: JOAN, KNOEL
;; APPLICANT: ROGAN, PETER
;; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATI
;; FILE REFERENCE: 30307
;; CURRENT APPLICATION NUMBER: US/09/573,080A
;; PRIOR FILING DATE: 2000-05-16
;; SOFTWARE: Patentin version 3.0
;; NUMBER OF SEQ ID NOS: 479
;; SEQ ID NO: 324
;; LENGTH: 230
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:

```
NAME/KEY: repeat region
LOCATION: (1)..(230)
OTHER INFORMATION: mer30
PUBLICATION INFORMATION:
AUTHORS: Jurka, J; Malchiewicz, J; Miosavljetic, A
TITLE: Prototypic sequences for human repetitive DNA
VOLUME: 35
ISSUE: 4
PAGES: 286-291
DATE: 1992-10-
DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
DATABASE ENTRY DATE:
DATABASE ENTRY DATE: 1996-01-26
US-09-573-080A-324
```

```
Query Match      84.0%; Score 21; DB 3; Length 230;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAACAATCT 25
Db 109 AAAAAAAAAATCGCAACAATCT 133
```

```
RESULT 8
US-09-270-767-8881
Sequence 8881, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
NUMBER OF SEQ ID NOS: 1999-03-17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8881
LENGTH: 488
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-8881
```

```
Query Match      83.2%; Score 20.8; DB 3; Length 488;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAACAATCT 24
Db 357 AAAAAAAAAATCGCAACAATCT 380
```

```
RESULT 9
US-09-270-767-24163
Sequence 24163, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
NUMBER OF SEQ ID NOS: 1999-03-17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24163
LENGTH: 488
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-24163
```

```
Query Match      83.2%; Score 20.8; DB 3; Length 488;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
```

```
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAACAATCT 24
Db 357 AAAAAAAAAATCGCAACAATCT 380
```

```
RESULT 10
US-09-949-016-105783
Sequence 105783, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: Venter, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 105783
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-105783
```

```
Query Match      83.2%; Score 20.8; DB 3; Length 601;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAACAATCT 24
Db 438 AAAAAAAAAATCGCAACAATCT 461
```

```
RESULT 11
US-09-843-472-5/C
Sequence 5, Application US/09843472
Patent No. 6544783
GENERAL INFORMATION:
APPLICANT: Perera, J. Ranjan
APPLICANT: Lu, Min
TITLE OF INVENTION: Polynucleotide Sequences from Rice
FILE REFERENCE: AKK-103G5XCI
CURRENT APPLICATION NUMBER: US/09/843,472
PRIOR FILING DATE: 2001-04-26
PRIOR APPLICATION NUMBER: 60/199,870
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/217,891
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: 60/218,366
PRIOR FILING DATE: 2000-07-13
PRIOR APPLICATION NUMBER: 60/227,231
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: 60/237,736
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/253,925
PRIOR FILING DATE: 2000-11-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 1834
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: misc_feature
```

```
; LOCATION: (1)..(1834)
; OTHER INFORMATION: n = a, c, g, or t.
US-09-843-472-5

Query Match
Best Local Similarity 83.2%; Score 20.8; DB 3; Length 1834;
Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATC 24
Db 1626 AAAAAAAAAATCGCAACCAATC 1603

RESULT 12
US-09-949-016-14688/c
; Sequence 14688, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14688
; LENGTH: 86857
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(86857)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14688

Query Match
Best Local Similarity 83.2%; Score 20.8; DB 3; Length 86857;
Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATC 24
Db 72008 AAAAAAAAAATCGCAACCAATC 71985

RESULT 13
US-09-949-016-105775/c
; Sequence 105775, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105775
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
```

```
; ORGANISM: Human
US-09-949-016-105775

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 601;
Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAAA 22
Db 72 AAAAAAAAAATCGCAACCAAA 51

RESULT 14
US-09-949-016-13495/c
; Sequence 13495, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13495
; LENGTH: 52655
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(52655)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13495

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 52655;
Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAAA 22
Db 656 AAAAAAAAAATCGCAACCAAA 635

RESULT 15
US-09-949-016-14688
; Sequence 14688, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14688
; LENGTH: 86857
; TYPE: DNA
; ORGANISM: Human
```

```
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(86857)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14688

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 86857;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAA 22
Db 63857 AAAAAAAAAATCGCAACAAA 63878

RESULT 16
US-09-949-016-14563
; Sequence 14563, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14563
; LENGTH: 92387
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(92387)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14563

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 92387;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAA 22
Db 47681 AAAAAAAAAATCGCAACAAA 47702

RESULT 17
US-09-949-016-14568
; Sequence 14568, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14568

FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(151295)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14569

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 151295;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAA 22
Db 47681 AAAAAAAAAATCGCAACAAA 47702

RESULT 18
US-09-949-016-14569
; Sequence 14569, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14569
; LENGTH: 151295
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(151295)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14569

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 151295;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAA 22
Db 47681 AAAAAAAAAATCGCAACAAA 47702

RESULT 19
US-09-949-016-14570
; Sequence 14570, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14570
```

```
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 14570
LENGTH: 151295
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(151295)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14570
```

```
Query Match          81.6% Score 20.4; DB 3; Length 151295;
Best Local Similarity 95.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
Db 47681 AAAAAAAAAATCGCAACAAA 47702
```

```
RESULT 20
US-09-949-016-14571
```

```
Sequence 14571, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 14571
LENGTH: 151295
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(151295)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14571
```

```
Query Match          81.6% Score 20.4; DB 3; Length 151295;
Best Local Similarity 95.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
Db 47681 AAAAAAAAAATCGCAACAAA 47702
```

```
RESULT 21
US-09-949-016-14572
```

```
Sequence 14572, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 14572
LENGTH: 151295
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(151295)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14572
```

```
Query Match          81.6% Score 20.4; DB 3; Length 151295;
Best Local Similarity 95.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
Db 47681 AAAAAAAAAATCGCAACAAA 47702
```

```
RESULT 22
US-09-949-016-17422/C
```

```
Sequence 17422, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 17422
LENGTH: 164061
TYPE: DNA
ORGANISM: Human
US-09-949-016-17422
```

```
Query Match          81.6% Score 20.4; DB 3; Length 164061;
Best Local Similarity 95.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
Db 135361 AAAAAAAAAATCGCAACAAA 135340
```

```
RESULT 23
US-09-949-016-14573
```

```
Sequence 14573, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
```

```

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14573
; LENGTH: 393753
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(393753)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14573

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 393753;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAACAAA 22
Db 290139 AAAAAAAAAATCGCAACAAA 290160

RESULT 24
US-09-949-016-14574
; Sequence 14574, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14574
; LENGTH: 393753
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(393753)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14574

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 393753;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAACAAA 22
Db 290139 AAAAAAAAAATCGCAACAAA 290160

RESULT 25
US-09-949-016-14546
; Sequence 14546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14546
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14546

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 818128;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAACAAA 22
Db 714514 AAAAAAAAAATCGCAACAAA 714535

RESULT 26
US-09-949-016-14547
; Sequence 14547, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14547
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14547

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 3; Length 818128;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAACAAA 22
Db 714514 AAAAAAAAAATCGCAACAAA 714535

RESULT 27
US-09-949-016-14548
; Sequence 14548, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
```

```

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14548
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14548
```

```

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAATCGCAACAAA 22
        |||
Db       714514 AAAAAAAAAATCGCAACAAA 714535
```

```

RESULT 28
US-09-949-016-14549
; Sequence 14549, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14549
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14549
```

```

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAATCGCAACAAA 22
        |||
Db       714514 AAAAAAAAAATCGCAACAAA 714535
```

```

RESULT 29
US-09-949-016-14550
; Sequence 14550, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14550
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14550
```

```

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAATCGCAACAAA 22
        |||
Db       714514 AAAAAAAAAATCGCAACAAA 714535
```

```

RESULT 30
US-09-949-016-14551
; Sequence 14551, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14551
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14551
```

```

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAATCGCAACAAA 22
        |||
Db       714514 AAAAAAAAAATCGCAACAAA 714535
```

```

RESULT 31
US-09-949-016-14552
; Sequence 14552, Application US/09949016
```

```
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14552
LENGTH: 818128
TYPE: DNA
ORGANISM: Human
NAME/KEY: misc_feature
LOCATION: (1)...(818128)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14552
```

```
Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
DB 714514 AAAAAAAAAATCGCAACAAA 714535
```

```
RESULT 32
US-09-949-016-14553
Sequence 14553, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14553
LENGTH: 818128
TYPE: DNA
ORGANISM: Human
NAME/KEY: misc_feature
LOCATION: (1)...(818128)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14553
```

```
Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
DB 714514 AAAAAAAAAATCGCAACAAA 714535
```

```
RESULT 33
US-09-949-016-14554
Sequence 14554, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14554
LENGTH: 818128
TYPE: DNA
ORGANISM: Human
NAME/KEY: misc_feature
LOCATION: (1)...(818128)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14554
```

```
Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
DB 714514 AAAAAAAAAATCGCAACAAA 714535
```

```
RESULT 34
US-09-949-016-14555
Sequence 14555, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14555
LENGTH: 818128
TYPE: DNA
ORGANISM: Human
NAME/KEY: misc_feature
LOCATION: (1)...(818128)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14555
```

```
Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCGCAACAAA 22
```


DB 714514 AAAAAAAAAATCGCAAAAAA 714535

RESULT 35
US-09-949-016-14556
; Sequence 14556, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14556
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14556

Query Match 81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAA 22
Db 714514 AAAAAAAAAATCGCAAAAAA 714535

RESULT 36
US-09-949-016-14557
; Sequence 14557, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14557
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14557

Query Match 81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAA 22
Db 714514 AAAAAAAAAATCGCAAAAAA 714535

RESULT 37
US-09-949-016-14558
; Sequence 14558, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14558
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14558

Query Match 81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAA 22
Db 714514 AAAAAAAAAATCGCAAAAAA 714535

RESULT 38
US-09-949-016-14559
; Sequence 14559, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14559
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14559

```
Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGCAACAAA 22
       |||
Db      714514 AAAAAAAAAATCGCAACAAA 714535

RESULT 39
US-09-949-016-14560
; Sequence 14560, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14560
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14560

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGCAACAAA 22
       |||
Db      714514 AAAAAAAAAATCGCAACAAA 714535

RESULT 40
US-09-949-016-14561
; Sequence 14561, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14561
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14561
```

```
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14561

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGCAACAAA 22
       |||
Db      714514 AAAAAAAAAATCGCAACAAA 714535

RESULT 41
US-09-949-016-14562
; Sequence 14562, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14562
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14562

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGCAACAAA 22
       |||
Db      714514 AAAAAAAAAATCGCAACAAA 714535

RESULT 42
US-09-949-016-14564
; Sequence 14564, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14564
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
```

```
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14564

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAA 22
Db 714514 AAAAAAAAAATCGCAACAA 714535

RESULT 43
US-09-949-016-14565
; Sequence 14565, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14565
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14565

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAA 22
Db 714514 AAAAAAAAAATCGCAACAA 714535

RESULT 44
US-09-949-016-14566
; Sequence 14566, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14566
```

```
LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14566

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAA 22
Db 714514 AAAAAAAAAATCGCAACAA 714535

RESULT 45
US-09-949-016-14567
; Sequence 14567, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14567
; LENGTH: 818128
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(818128)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14567

Query Match      81.6%; Score 20.4; DB 3; Length 818128;
Best Local Similarity 95.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAA 22
Db 714514 AAAAAAAAAATCGCAACAA 714535

RESULT 46
US-09-949-016-43146/C
; Sequence 43146, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 43146
 LENGTH: 601
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-43146

Query Match
 Best Local Similarity 80.8%; Score 20.2; DB 3; Length 601;
 Pred. No. 2.8e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
 |||||
 Db 294 AAAAAAAAAATCGCAACCAATAT 270

RESULT 47
 US-09-949-016-43375/c
 Sequence 43375, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CLO01307
 CURRENT APPLICATION NUMBER: US/09/949,016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 43375
 LENGTH: 601
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-43375

Query Match
 Best Local Similarity 80.8%; Score 20.2; DB 3; Length 601;
 Pred. No. 2.8e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
 |||||
 Db 294 AAAAAAAAAATCGCAACCAATAT 270

RESULT 48
 US-09-949-016-43604/c
 Sequence 43604, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CLO01307
 CURRENT APPLICATION NUMBER: US/09/949,016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 43604
 LENGTH: 601
 TYPE: DNA

ORGANISM: Human
 US-09-949-016-43604

Query Match
 Best Local Similarity 80.8%; Score 20.2; DB 3; Length 601;
 Pred. No. 2.8e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
 |||||
 Db 294 AAAAAAAAAATCGCAACCAATAT 270

RESULT 49
 US-09-949-016-64297
 Sequence 64297, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CLO01307
 CURRENT APPLICATION NUMBER: US/09/949,016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 64297
 LENGTH: 601
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-64297

Query Match
 Best Local Similarity 80.8%; Score 20.2; DB 3; Length 601;
 Pred. No. 2.8e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
 |||||
 Db 304 AAAAAAAAAATCGCAACCAATCT 328

RESULT 50
 US-09-949-016-78412
 Sequence 78412, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CLO01307
 CURRENT APPLICATION NUMBER: US/09/949,016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 78412
 LENGTH: 601
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-78412

Query Match
 Best Local Similarity 80.8%; Score 20.2; DB 3; Length 601;
 Pred. No. 2.8e+02;

Matches	22;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
Qy	1	AAAAAAAAAATCGCAACAAATCT	25						
Db	414	AAAAAAAAAATCCCAAAAAATCT	438						

Search completed: December 14, 2005, 07:44:03
Job time : 79.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds
(without alignments)
800.703 Million cell updates/sec

Title: US-10-681-773-2

Perfect score: 25

Sequence: 1 aaaaaaaaaatcgagacaatc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/pdata/1/ina/1_COMB.seq:*
- 2: /cgn2_6/pdata/1/ina/5_COMB.seq:*
- 3: /cgn2_6/pdata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/pdata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/pdata/1/ina/H_COMB.seq:*
- 6: /cgn2_6/pdata/1/ina/PCrUS_COMB.seq:*
- 7: /cgn2_6/pdata/1/ina/PP_COMB.seq:*
- 8: /cgn2_6/pdata/1/ina/RE_COMB.seq:*
- 9: /cgn2_6/pdata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.8	87.2	601	US-09-949-016-128549	Sequence 128549, A
2	21.8	87.2	20935	US-09-949-016-15383	Sequence 15383, A
3	21.8	87.2	40168	US-09-949-016-13225	Sequence 13225, A
4	21.8	87.2	46841	US-09-949-016-13466	Sequence 13466, A
5	21.8	87.2	60141	US-09-949-016-15874	Sequence 15874, A
6	21.8	87.2	360470	US-09-949-016-11173	Sequence 13173, A
7	20.2	80.8	601	US-09-949-016-182939	Sequence 182939, A
8	20.2	80.8	601	US-09-949-016-182940	Sequence 182940, A
9	20.2	80.8	601	US-09-949-016-182941	Sequence 182941, A
10	20.2	80.8	41737	US-09-949-016-12204	Sequence 12204, A
11	20.2	80.8	41741	US-09-949-016-15983	Sequence 15983, A
12	20.2	80.8	71251	US-09-949-016-15332	Sequence 15332, A
13	20.2	80.8	91772	US-09-949-016-15568	Sequence 15568, A
14	19.8	79.2	230	US-09-573-080A-124	Sequence 324, App
15	19.8	79.2	601	US-09-949-016-78415	Sequence 78415, A
16	19.8	79.2	601	US-09-949-016-178188	Sequence 178188, A
17	19.8	79.2	601	US-09-949-016-178189	Sequence 178189, A
18	19.8	79.2	5596	US-09-078-294-5	Sequence 5, Appl
19	19.8	79.2	22121	US-09-949-016-14359	Sequence 14359, A
20	19.8	79.2	22121	US-09-949-016-14909	Sequence 14909, A
21	19.8	79.2	33061	US-09-949-016-16848	Sequence 16848, A
22	19.8	79.2	80246	US-09-078-294-4	Sequence 4, Appl
23	19.8	79.2	80595	US-09-078-294-3	Sequence 3, Appl
24	19.2	76.8	444	US-09-270-767-4829	Sequence 4829, Ap

C	25	19.2	76.8	444	3	US-09-270-767-20111
C	26	19.2	76.8	488	3	US-09-270-767-8881
C	27	19.2	76.8	488	3	US-09-270-767-24163
C	28	19.2	76.8	498	3	US-09-134-001C-2267
C	29	19.2	76.8	601	3	US-09-949-016-105783
C	30	19.2	76.8	601	3	US-09-949-016-144585
C	31	19.2	76.8	1834	3	US-09-843-472-5
C	32	19.2	76.8	77618	3	US-09-949-016-11768
C	33	19.2	76.8	85369	3	US-09-949-016-12171
C	34	19.2	76.8	86857	3	US-09-949-016-14688
C	35	19.2	76.8	121982	3	US-09-949-016-12085
C	36	19.2	76.8	121982	3	US-09-949-016-14105
C	37	19.2	76.8	162025	3	US-09-834-700-13
C	38	19.2	76.8	162025	3	US-09-834-700-14
C	39	19.2	76.8	162025	3	US-09-834-700-17
C	40	19.2	76.8	162025	3	US-09-834-700-18
C	41	19.2	76.8	192506	3	US-09-949-016-15830
C	42	18.8	75.2	106	3	US-09-513-999C-33251
C	43	18.8	75.2	601	3	US-09-949-016-58039
C	44	18.8	75.2	601	3	US-09-949-016-105775
C	45	18.8	75.2	601	3	US-09-949-016-189946
C	46	18.8	75.2	601	3	US-09-949-016-201980
C	47	18.8	75.2	601	3	US-09-949-016-201981
C	48	18.8	75.2	1152	3	US-09-270-767-6587
C	49	18.8	75.2	1152	3	US-09-270-767-21869
C	50	18.8	75.2	3585	3	US-09-270-767-14479
C	51	18.8	75.2	14255	3	US-09-949-016-16875
C	52	18.8	75.2	15593	3	US-09-949-016-17177
C	53	18.8	75.2	32244	3	US-09-949-016-16806
C	54	18.8	75.2	52655	3	US-09-949-016-13495
C	55	18.8	75.2	88850	3	US-09-949-016-13424
C	56	18.8	75.2	86887	3	US-09-949-016-14568
C	57	18.8	75.2	92387	3	US-09-949-016-14563
C	58	18.8	75.2	96845	3	US-09-949-016-13658
C	59	18.8	75.2	137753	3	US-09-949-016-17404
C	60	18.8	75.2	144158	3	US-09-949-016-11755
C	61	18.8	75.2	144158	3	US-09-949-016-12936
C	62	18.8	75.2	151295	3	US-09-949-016-14568
C	63	18.8	75.2	151295	3	US-09-949-016-14570
C	64	18.8	75.2	151295	3	US-09-949-016-14571
C	65	18.8	75.2	151295	3	US-09-949-016-14572
C	66	18.8	75.2	164061	3	US-09-949-016-17422
C	67	18.8	75.2	317366	3	US-09-949-016-16001
C	68	18.8	75.2	393753	3	US-09-949-016-14573
C	69	18.8	75.2	393753	3	US-09-949-016-14574
C	70	18.8	75.2	818128	3	US-09-949-016-14546
C	71	18.8	75.2	818128	3	US-09-949-016-14547
C	72	18.8	75.2	818128	3	US-09-949-016-14548
C	73	18.8	75.2	818128	3	US-09-949-016-14549
C	74	18.8	75.2	818128	3	US-09-949-016-14550
C	75	18.8	75.2	818128	3	US-09-949-016-14551
C	76	18.8	75.2	818128	3	US-09-949-016-14552
C	77	18.8	75.2	818128	3	US-09-949-016-14553
C	78	18.8	75.2	818128	3	US-09-949-016-14554
C	79	18.8	75.2	818128	3	US-09-949-016-14555
C	80	18.8	75.2	818128	3	US-09-949-016-14556
C	81	18.8	75.2	818128	3	US-09-949-016-14557
C	82	18.8	75.2	818128	3	US-09-949-016-14558
C	83	18.8	75.2	818128	3	US-09-949-016-14559
C	84	18.8	75.2	818128	3	US-09-949-016-14560
C	85	18.8	75.2	818128	3	US-09-949-016-14561
C	86	18.8	75.2	818128	3	US-09-949-016-14562
C	87	18.8	75.2	818128	3	US-09-949-016-14564
C	88	18.8	75.2	818128	3	US-09-949-016-14565
C	89	18.8	75.2	818128	3	US-09-949-016-14566
C	90	18.8	75.2	818128	3	US-09-949-016-14567
C	91	18.8	75.2	818128	3	US-09-949-016-14567
C	92	18.6	74.4	88	3	US-09-513-999C-36378
C	93	18.6	74.4	83	3	US-09-513-999C-36145
C	94	18.6	74.4	195	3	US-09-248-796A-11360
C	95	18.6	74.4	601	3	US-09-949-016-22236
C	96	18.6	74.4	601	3	US-09-949-016-11565
C	97	18.6	74.4	601	3	US-09-949-016-14566

Sequence 20111, A	Sequence 8881, Ap	Sequence 24163, A	Sequence 2267, Ap	Sequence 105783, A	Sequence 144585, A	Sequence 5, Appl	Sequence 11768, A	Sequence 12171, A	Sequence 14688, A	Sequence 12085, A	Sequence 14105, A	Sequence 13, Appl	Sequence 17, Appl	Sequence 18, Appl	Sequence 15830, A	Sequence 33251, A	Sequence 58039, A	Sequence 105775, A	Sequence 189946, A	Sequence 201980, A	Sequence 201981, A	Sequence 6587, Ap	Sequence 21659, A	Sequence 14479, A	Sequence 16875, A	Sequence 17177, A	Sequence 16806, A	Sequence 13495, A	Sequence 13424, A	Sequence 14688, A	Sequence 14568, A	Sequence 14569, A	Sequence 14570, A	Sequence 14571, A	Sequence 14572, A	Sequence 17422, A	Sequence 16001, A	Sequence 14573, A	Sequence 14574, A	Sequence 14546, A	Sequence 14547, A	Sequence 14548, A	Sequence 14549, A	Sequence 14550, A	Sequence 14551, A	Sequence 14552, A	Sequence 14553, A	Sequence 14554, A	Sequence 14555, A	Sequence 14556, A	Sequence 14557, A	Sequence 14558, A	Sequence 14559, A	Sequence 14560, A	Sequence 14561, A	Sequence 14562, A	Sequence 14564, A	Sequence 14565, A	Sequence 14566, A	Sequence 14567, A	Sequence 36145, A	Sequence 36378, A	Sequence 11360, A	Sequence 32236, A	Sequence 41565, A	Sequence 41566, A
-------------------	-------------------	-------------------	-------------------	--------------------	--------------------	------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	--------------------	--------------------	--------------------	--------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

```
98 18.6 74.4 601 3 US-09-949-016-41567 Sequence 41567, A
99 18.6 74.4 601 3 US-09-949-016-41568 Sequence 41568, A
100 18.6 74.4 601 3 US-09-949-016-41569 Sequence 41569, A
101 18.6 74.4 601 3 US-09-949-016-41570 Sequence 41570, A
102 18.6 74.4 601 3 US-09-949-016-41589 Sequence 41589, A
103 18.6 74.4 601 3 US-09-949-016-43146 Sequence 43146, A
104 18.6 74.4 601 3 US-09-949-016-43375 Sequence 43375, A
105 18.6 74.4 601 3 US-09-949-016-43604 Sequence 43604, A
106 18.6 74.4 601 3 US-09-949-016-50415 Sequence 50415, A
107 18.6 74.4 601 3 US-09-949-016-50417 Sequence 50417, A
108 18.6 74.4 601 3 US-09-949-016-51167 Sequence 51167, A
109 18.6 74.4 601 3 US-09-949-016-61152 Sequence 61152, A
110 18.6 74.4 601 3 US-09-949-016-64297 Sequence 64297, A
111 18.6 74.4 601 3 US-09-949-016-66332 Sequence 66332, A
112 18.6 74.4 601 3 US-09-949-016-70427 Sequence 70427, A
113 18.6 74.4 601 3 US-09-949-016-70428 Sequence 70428, A
114 18.6 74.4 601 3 US-09-949-016-78412 Sequence 78412, A
115 18.6 74.4 601 3 US-09-949-016-78413 Sequence 78413, A
116 18.6 74.4 601 3 US-09-949-016-78416 Sequence 78416, A
117 18.6 74.4 601 3 US-09-949-016-87942 Sequence 87942, A
118 18.6 74.4 601 3 US-09-949-016-87942 Sequence 87942, A
119 18.6 74.4 601 3 US-09-949-016-11136 Sequence 11136, A
120 18.6 74.4 601 3 US-09-949-016-126006 Sequence 126006, A
121 18.6 74.4 601 3 US-09-949-016-126007 Sequence 126007, A
122 18.6 74.4 601 3 US-09-949-016-150382 Sequence 150382, A
123 18.6 74.4 601 3 US-09-949-016-150383 Sequence 150383, A
124 18.6 74.4 601 3 US-09-949-016-168632 Sequence 168632, A
125 18.6 74.4 601 3 US-09-949-016-190379 Sequence 190379, A
126 18.6 74.4 601 3 US-09-949-016-192634 Sequence 192634, A
127 18.6 74.4 601 3 US-09-949-016-194232 Sequence 194232, A
128 18.6 74.4 601 3 US-09-949-016-194233 Sequence 194233, A
129 18.6 74.4 601 3 US-09-949-016-194234 Sequence 194234, A
130 18.6 74.4 601 3 US-09-949-016-194235 Sequence 194235, A
131 18.6 74.4 601 3 US-09-949-016-194236 Sequence 194236, A
132 18.6 74.4 601 3 US-09-949-016-194237 Sequence 194237, A
133 18.6 74.4 601 3 US-09-949-016-194238 Sequence 194238, A
134 18.6 74.4 601 3 US-09-949-002-1984 Sequence 1984, A
135 18.6 74.4 601 3 US-09-949-002-1985 Sequence 1985, A
136 18.6 74.4 601 3 US-09-949-002-9584 Sequence 9584, A
137 18.6 74.4 601 3 US-09-949-002-9585 Sequence 9585, A
138 18.6 74.4 601 3 US-09-270-767-2247 Sequence 2247, A
139 18.6 74.4 601 3 US-09-270-767-17529 Sequence 17529, A
140 18.6 74.4 601 3 US-09-949-016-19511 Sequence 19511, A
141 18.6 74.4 601 3 US-09-949-016-732 Sequence 732, A
142 18.6 74.4 601 3 US-09-949-016-4753 Sequence 4753, A
143 18.6 74.4 601 3 US-10-131-827-8854 Sequence 8854, A
144 18.6 74.4 601 3 US-09-949-016-15353 Sequence 15353, A
145 18.6 74.4 601 3 US-09-949-002-586 Sequence 586, A
146 18.6 74.4 601 3 US-09-949-002-844 Sequence 844, A
147 18.6 74.4 601 3 US-09-210-748A-3 Sequence 3, A
148 18.6 74.4 601 3 US-09-939-581A-3 Sequence 3, A
149 18.6 74.4 601 3 US-09-771-357-102 Sequence 102, A
150 18.6 74.4 601 3 US-10-059-579A-102 Sequence 102, A
```

ALIGNMENTS

```
RESULT 1
US-09-949-016-128549/C
; Sequence 128549, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 128549
;; LENGTH: 601
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-128549
Query Match 87.2%; Score 21.8; DB 3; Length 601;
Best Local Similarity 92.0%; Pred. No. 33;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 519 AAAAAAAAAACCGCAGCAAAATAT 495

RESULT 2
US-09-949-016-15383
; Sequence 15383, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15383
; LENGTH: 20935
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15383
Query Match 87.2%; Score 21.8; DB 3; Length 20935;
Best Local Similarity 92.0%; Pred. No. 39;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 6497 AAAAAAAAAACCGCAGCAAAATAT 6521

RESULT 3
US-09-949-016-13225/C
; Sequence 13225, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13225
```



```

; LENGTH: 40168
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1) ..(40168)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13225

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 3; Length 40168;
Best Local Similarity 92.0%; Pred. No. 40;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 8976 AAAAAAAAAATCGCAGCAATCT 8952

RESULT 4
US-09-949-016-13466/C
; Sequence 13466, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13466
; LENGTH: 46841
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13466

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 3; Length 46841;
Best Local Similarity 92.0%; Pred. No. 41;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 20040 AAAAAAAAAATCGCAGCAATCT 20016

RESULT 5
US-09-949-016-15874/C
; Sequence 15874, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15874
; LENGTH: 60141
; TYPE: DNA

; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1) ..(60141)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15874

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 3; Length 60141;
Best Local Similarity 92.0%; Pred. No. 41;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 47220 AAAAAAAAAATCGCAGCAATCT 47196

RESULT 6
US-09-949-016-13173/C
; Sequence 13173, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13173
; LENGTH: 360470
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13173

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 3; Length 360470;
Best Local Similarity 92.0%; Pred. No. 45;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 301343 AAAAAAAAAATCGCAGCAATCT 301319

RESULT 7
US-09-949-016-18293/C
; Sequence 18293, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18293
; LENGTH: 601
; TYPE: DNA
```

ORGANISM: Human
US-09-949-016-182939

Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAGCAATCT 25
Db 424 AAAAAAAAAACACAGCAATCT 400

RESULT 8
US-09-949-016-182940/C
Sequence 182940, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CLO01307

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 182940

LENGTH: 601

TYPE: DNA

ORGANISM: Human

US-09-949-016-182940

Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAGCAATCT 25
Db 540 AAAAAAAAAACACAGCAATCT 516

RESULT 9

US-09-949-016-182941/C

Sequence 182941, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CLO01307

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 182941

LENGTH: 601

TYPE: DNA

ORGANISM: Human

US-09-949-016-182941

Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAGCAATCT 25
Db 586 AAAAAAAAAACACAGCAATCT 562

RESULT 10
US-09-949-016-12204
Sequence 12204, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CLO01307

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 12204

LENGTH: 41737

TYPE: DNA

ORGANISM: Human

FEATURES:

NAME/KEY: misc_feature

LOCATION: (1)...(41737)

OTHER INFORMATION: n = A,T,C or G

Query Match 80.8%; Score 20.2; DB 3; Length 41737;
Best Local Similarity 88.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAGCAATCT 25
Db 11861 AAAAAAAAAACACAGCAATCT 11885

RESULT 11

US-09-949-016-16983

Sequence 16983, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CLO01307

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 16983

LENGTH: 41741

TYPE: DNA

ORGANISM: Human

FEATURES:

NAME/KEY: misc_feature

LOCATION: (1)...(41741)

OTHER INFORMATION: n = A,T,C or G

US-09-949-016-16983

```
Query Match      80.8%; Score 20.2; DB 3; Length 41741;
Best Local Similarity 88.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 11861 AAAAAAAAAACACAGAGAAATCT 11885

RESULT 12
US-09-949-016-15332
; Sequence 15332, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15332
; LENGTH: 71251
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15332

Query Match      80.8%; Score 20.2; DB 3; Length 71251;
Best Local Similarity 88.0%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 50265 AAAAAAAAAATCACAATAAAATCT 50289

RESULT 13
US-09-949-016-15568
; Sequence 15568, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15568
; LENGTH: 91772
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(91772)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15568

Query Match      80.8%; Score 20.2; DB 3; Length 91772;
Best Local Similarity 88.0%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 48082 AAAAAAAAAATCACAATAAAATCT 48106

RESULT 14
US-09-573-080A-324
; Sequence 324, Application US/09573080A
; Patent No. 6828097
; GENERAL INFORMATION:
; APPLICANT: JOAN, KNOEL
; APPLICANT: ROGAN, PETER
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATI
; FILE REFERENCE: 30307
; CURRENT APPLICATION NUMBER: US/09/573,080A
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 479
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 324
; LENGTH: 230
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: (1)..(230)
; OTHER INFORMATION: mer30
; PUBLICATION INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Jurka, J; Malichiewicz, J; Miosavljivic, A
; TITLE: Prototypic sequences for human repetitive DNA
; JOURNAL: Journal of Molecular Evolution
; VOLUME: 35
; ISSUE: 4
; PAGES: 286-291
; DATE: 1992-10-
; DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
; DATABASE ENTRY DATE:
; DATABASE ENTRY DATE: 1996-01-26
US-09-573-080A-324

Query Match      79.2%; Score 19.8; DB 3; Length 230;
Best Local Similarity 84.0%; Pred. No. 1.7e+02;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 109 AAAAAAAAAATCCCAAAAATCT 133

RESULT 15
US-09-949-016-78415
; Sequence 78415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78415
```

LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-78415

Query Match 79.2%; Score 19.8; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACCAATCT 25
DB 285 AAAAAAAAAATCGACGACCAATCT 309

RESULT 16
US-09-949-016-178188
Sequence 178188, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT FILING DATE: 2000-04-14
PRIOR FILING DATE: 2000-04-14
PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-10-03
PRIOR FILING DATE: 2000-10-03
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 178188
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-178188

Query Match 79.2%; Score 19.8; DB 3; Length 601;
Best Local Similarity 91.3%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACCAAT 23
DB 153 AAAAAAAAAATCGACGACCAAT 175

RESULT 17
US-09-949-016-178189
Sequence 178189, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT FILING DATE: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR FILING DATE: 2000-04-14
PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-10-03
PRIOR FILING DATE: 2000-10-03
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 178189
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-178189

Query Match 79.2%; Score 19.8; DB 3; Length 601;
Best Local Similarity 91.3%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACCAAT 23
DB 145 AAAAAAAAAATCGACGACCAAT 167

RESULT 18
US-09-078-294-5/c
Sequence 5, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT FILING DATE: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 5
LENGTH: 5596
TYPE: DNA
ORGANISM: BAC-F2 contig 1
US-09-078-294-5

Query Match 79.2%; Score 19.8; DB 3; Length 5596;
Best Local Similarity 91.3%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCGACGACCAATC 24
DB 711 AAAAAAAAAATCAAGACCAATC 689

RESULT 19
US-09-949-016-14359/c
Sequence 14359, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT FILING DATE: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-10-03
PRIOR FILING DATE: 2000-10-03
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 14359
LENGTH: 22121
TYPE: DNA
ORGANISM: Human
US-09-949-016-14359

Query Match 79.2%; Score 19.8; DB 3; Length 22121;
Best Local Similarity 91.3%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACCAAT 23
DB 20248 AAAAAAAAAATCGACGACCAAT 20226

RESULT 20

```
US-09-949-016-14909/C
; Sequence 14909, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14909
; LENGTH: 22121
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14909

Query Match          79.2%; Score 19.8; DB 3; Length 22121;
Best Local Similarity 91.3%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACCAAT 23
Db 20248 AAAAAAAAAATCTGAGAAAAAT 20226

RESULT 21
US-09-949-016-16848
; Sequence 16848, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16848
; LENGTH: 33061
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16848

Query Match          79.2%; Score 19.8; DB 3; Length 33061;
Best Local Similarity 91.3%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACCAAT 23
Db 15127 AAAAAAAAAAGCAGACCAAT 15149

RESULT 22
US-09-078-294-4/C
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match          79.2%; Score 19.8; DB 3; Length 80246;
Best Local Similarity 91.3%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCGACGACCAATC 24
Db 68701 AAAAAAAAAATCAAGACCAATC 68679

RESULT 23
US-09-078-294-3/C
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match          79.2%; Score 19.8; DB 3; Length 80595;
Best Local Similarity 91.3%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCGACGACCAATC 24
Db 68964 AAAAAAAAAATCAAGACCAATC 68942

RESULT 24
US-09-270-767-4829/C
; Sequence 4829, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7336-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4829
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: n means any nucleotide
US-09-270-767-4829

Query Match          76.8%; Score 19.2; DB 3; Length 444;
```

Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGCAATCT 25
Db 383 AAAAAAAAAACGTAACCAATC 359

RESULT 25

US-09-270-767-20111/C
; Sequence 20111, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 20111
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: n means any nucleotide
US-09-270-767-20111

Query Match 76.8%; Score 19.2; DB 3; Length 444;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGCAATCT 25
Db 383 AAAAAAAAAACGTAACCAATCT 359

RESULT 26

US-09-270-767-8881
; Sequence 8881, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8881
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-8881

Query Match 76.8%; Score 19.2; DB 3; Length 488;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGCAATCT 24
Db 357 AAAAAAAAAACGTAACCAATC 380

RESULT 27

US-09-270-767-24163
; Sequence 24163, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24163
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24163

Query Match 76.8%; Score 19.2; DB 3; Length 488;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGCAATCT 24
Db 357 AAAAAAAAAACGTAACCAATC 380

RESULT 28

US-09-134-001C-2267
; Sequence 2267, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2267
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2267

Query Match 76.8%; Score 19.2; DB 3; Length 498;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGACGCAATCT 25
Db 435 AAAAAAAAAACGTAACCAATCT 458

RESULT 29

US-09-949-016-105783
; Sequence 105783, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CI001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 105783
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-105783

```
Query Match          76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGACAATC 24
    |||||
Db 438 AAAAAAAAAAACCAACAATC 461

RESULT 30
US-09-949-016-144585/c
; Sequence 144585, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144585
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-144585

Query Match          76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGAGACAATCT 25
    |||||
Db 399 AGAAAAATAATCAGACAAATCT 376

RESULT 31
US-09-843-472-5/c
; Sequence 5, Application US/09843472
; Patent No. 6544783
; GENERAL INFORMATION:
; APPLICANT: Perera, J. Ranjan
; APPLICANT: Lu, Min
; APPLICANT: Ray, Animesh
; TITLE OF INVENTION: Polynucleotide Sequences from Rice
; FILE REFERENCE: AKK-103C5XC1
; CURRENT APPLICATION NUMBER: US/09/843,472
; CURRENT FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: 60/199,870
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/217,891
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/218,366
; PRIOR FILING DATE: 2000-07-13
; PRIOR APPLICATION NUMBER: 60/227,231
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/237,736
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/253,925
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1834
; TYPE: DNA
```

```
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(1834)
; OTHER INFORMATION: n = a, c, g, or t.
US-09-843-472-5

Query Match          76.8%; Score 19.2; DB 3; Length 1834;
Best Local Similarity 87.5%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGACAATC 24
    |||||
Db 1626 AAAAAAAAAATCACAACAATC 1603

RESULT 32
US-09-949-016-11768
; Sequence 11768, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11768
; LENGTH: 77618
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(77618)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11768

Query Match          76.8%; Score 19.2; DB 3; Length 77618;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGACAATC 24
    |||||
Db 57996 AAAAAAAAAATCGATCCAATC 58019

RESULT 33
US-09-949-016-12171
; Sequence 12171, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 12171
; LENGTH: 85369
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(85369)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12171
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 85369;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAGCAATC 24
Db 24112 AAAAAAAAAAGCAGCAATTC 24135
```

```
RESULT 34
US-09-949-016-14688/C
; Sequence 14688, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14688
; LENGTH: 86857
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(86857)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14688
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 86857;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAGCAATC 24
Db 72008 AAAAAAAAAACCAACAATC 71985
```

```
RESULT 35
US-09-949-016-12085/C
; Sequence 12085, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12085
; LENGTH: 121982
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12085
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 121982;
Best Local Similarity 87.5%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAGCAATC 24
Db 80426 AAAAAAAAAAGCAGCAATC 80403
```

```
RESULT 36
US-09-949-016-14105/C
; Sequence 14105, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14105
; LENGTH: 121982
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14105
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 121982;
Best Local Similarity 87.5%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAGCAATC 24
Db 80426 AAAAAAAAAAGCAGCAATC 80403
```

```
RESULT 37
US-09-834-700-13/C
; Sequence 13, Application US/09834700
; Patent No. 6958214
; GENERAL INFORMATION:
; APPLICANT: Braun, A.
; TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
; FILE REFERENCE: 24736-2035
; CURRENT APPLICATION NUMBER: US/09/834,700
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/217,251
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: 60/240,335
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 162025
; TYPE: DNA
; ORGANISM: Homo Sapien
```



```
FEATURE:
; NAME/KEY: allele
; LOCATION: 83587
; OTHER INFORMATION: Nucleotide sequence of Chromosome 17 with genomic
; OTHER INFORMATION: sequence of the allelic variant AKAP10-6
US-09-834-700-13

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 3; Length 162025;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGACAAATC 24
Db 12499 AAAAAAAAAATTCAGACATATC 12476

RESULT 38
US-09-834-700-14/c
; Sequence 14, Application US/09834700
; Patent No. 6958214
; GENERAL INFORMATION:
; APPLICANT: Braun, A.
; TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
; FILE REFERENCE: 24736-2035
; CURRENT APPLICATION NUMBER: US/09/834,700
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/217,251
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: 60/240,335
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 162025
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: allele
; LOCATION: 129600
; OTHER INFORMATION: Nucleotide sequence of chromosome 17 containing
; OTHER INFORMATION: the genomic sequence of the allelic variant
US-09-834-700-14

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 3; Length 162025;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGACAAATC 24
Db 12499 AAAAAAAAAATTCAGACATATC 12476

RESULT 39
US-09-834-700-17/c
; Sequence 17, Application US/09834700
; Patent No. 6958214
; GENERAL INFORMATION:
; APPLICANT: Braun, A.
; TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
; FILE REFERENCE: 24736-2035
; CURRENT APPLICATION NUMBER: US/09/834,700
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/217,251
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: 60/240,335
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 162025
```

```
TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Nucleotide sequence of chromosome 17 containing
; OTHER INFORMATION: the genomic sequence of the predominate allele of
US-09-834-700-17

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 3; Length 162025;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGACAAATC 24
Db 12499 AAAAAAAAAATTCAGACATATC 12476

RESULT 40
US-09-834-700-18/c
; Sequence 18, Application US/09834700
; Patent No. 6958214
; GENERAL INFORMATION:
; APPLICANT: Braun, A.
; TITLE OF INVENTION: POLYMORPHIC KINASE ANCHOR PROTEINS AND
; FILE REFERENCE: 24736-2035
; CURRENT APPLICATION NUMBER: US/09/834,700
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/217,251
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: 60/240,335
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 162025
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: allele
; LOCATION: 156,277
; OTHER INFORMATION: Nucleotide sequence of chromosome 17 containing
; OTHER INFORMATION: the genomic sequence of the allelic variant
US-09-834-700-18

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 3; Length 162025;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGACAAATC 24
Db 12499 AAAAAAAAAATTCAGACATATC 12476

RESULT 41
US-09-949-016-15830/c
; Sequence 15830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```

PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15830
LENGTH: 192506
TYPE: DNA
ORGANISM: Human
US-09-949-016-15830

Query Match 76.8%; Score 19.2; DB 3; Length 192506;
Best Local Similarity 87.5%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGACGACAAATCT 25
Db 60416 AAAAAAAAAATCGACGACAAATCT 60393

RESULT 42
US-09-513-999C-33251/c
Sequence 33251, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclet, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59, US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36661
SOFTWARE: Patent.pm
SEQ ID NO 33251
LENGTH: 106
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12
OTHER INFORMATION: r=a or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: 47
OTHER INFORMATION: w=a or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 55
OTHER INFORMATION: k=g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 71
OTHER INFORMATION: k=g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 82
OTHER INFORMATION: y=c or t
US-09-513-999C-33251

Query Match 75.2%; Score 18.8; DB 3; Length 106;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGACAAATC 24
Db 95 AAAAAAAAAATCGACGACAAATC 72

RESULT 43
US-09-949-016-58039/c
Sequence 58039, Application US/09949016

Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/231,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 58039
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-58039

Query Match 75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGACAAA 22
Db 232 AAAAAAAAAATCGACGACAAA 211

RESULT 44
US-09-949-016-105775/c
Sequence 105775, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 105775
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-105775

Query Match 75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGACAAA 22
Db 72 AAAAAAAAAATCGACGACAAA 51

RESULT 45
US-09-949-016-189946/c
Sequence 189946, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 189946
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-189946

```
FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 189946
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-189946

Query Match          75.2%  Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGACACAAA 22
   |||||
Db 295 AAAAAAAAAATCGACACAAA 274

RESULT 46
US-09-949-016-201980/c
; Sequence 201980, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 201980
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-201980

Query Match          75.2%  Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGACACAAA 22
   |||||
Db 301 AAAAAAAAAATCGACACAAA 280

RESULT 47
US-09-949-016-201981/c
; Sequence 201981, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 201981
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-201981

Query Match          75.2%  Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGACACAAA 22
   |||||
Db 211 AAAAAAAAAATCGACACAAA 190

RESULT 48
US-09-270-767-6587/c
; Sequence 6587, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 6587
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-6587

Query Match          75.2%  Score 18.8; DB 3; Length 1152;
Best Local Similarity 90.9%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGACACAAA 22
   |||||
Db 506 AAAAAAAAAATCGACACAAA 485

RESULT 49
US-09-270-767-21869/c
; Sequence 21869, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 21869
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-21869

Query Match          75.2%  Score 18.8; DB 3; Length 1152;
Best Local Similarity 90.9%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGACACAAA 22
   |||||
Db 506 AAAAAAAAAATCGACACAAA 485
```

RESULT 50
 US-09-270-767-14479
 ; Sequence 14479, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; NUMBER OF SEQ ID NOS: 1999-03-17
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 14479
 ; LENGTH: 3585
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-14479

Query Match 75.2%; Score 18.8; DB 3; Length 3585;
 Best Local Similarity 90.9%; Pred. No. 4.4e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGACAAA 22
 |||||
 Db 2029 AAAAAAAAAACCGCAGAGAAA 2050

Search completed: December 14, 2005, 07:44:15
 Job time : 67.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds
(without alignments)
800.703 Million cell updates/sec

Title: US-10-681-773-3

Perfect score: 25

Sequence: 1 aaaaaaactaaagctgacctc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCRTUS_COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
- 8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
- 9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.8	83.2	93510	3	US-09-949-016-15095
C 2	20.2	80.8	627	3	US-09-270-767-9667
C 3	20.2	80.8	627	3	US-09-270-767-24949
C 4	19.8	79.2	945	3	US-09-543-681A-4161
C 5	19.8	79.2	1184	3	US-09-270-767-9472
C 6	19.8	79.2	1184	3	US-09-270-767-24754
C 7	19.2	76.8	50	3	US-08-851-843A-102
C 8	19.2	76.8	50	3	US-08-974-459A-587
C 9	19.2	76.8	50	3	US-08-854-050-102
C 10	19.2	76.8	50	3	US-09-430-323-102
C 11	19.2	76.8	50	3	US-09-402-181B-587
C 12	19.2	76.8	50	3	US-09-721-456-587
C 13	19.2	76.8	50	3	US-09-766-251-102
C 14	19.2	76.8	50	3	US-10-054-295-102
C 15	19.2	76.8	50	3	US-09-438-486A-102
C 16	19.2	76.8	52	3	US-09-057-328-11
C 17	19.2	76.8	52	3	US-09-256-373-11
C 18	19.2	76.8	52	3	US-09-091-590A-15
C 19	19.2	76.8	52	3	US-09-228-455-1
C 20	19.2	76.8	52	3	US-09-228-455-2
C 21	19.2	76.8	52	3	US-09-276-533A-17
C 22	19.2	76.8	52	3	US-10-158-735-17
C 23	19.2	76.8	52	3	US-09-443-282B-12
C 24	19.2	76.8	52	3	US-10-154-517-1

C 25	19.2	76.8	52	3	US-10-154-517-2	Sequence 2, Appli
C 26	19.2	76.8	304	3	US-09-621-976-17988	Sequence 17988, A
C 27	19.2	76.8	601	3	US-09-949-016-52788	Sequence 52788, A
C 28	19.2	76.8	669	3	US-09-328-111-342	Sequence 342, App
C 29	19.2	76.8	9139	3	US-09-322-478-22	Sequence 22, Appl
C 30	19.2	76.8	9139	3	US-09-586-106D-22	Sequence 22, Appl
C 31	19.2	76.8	9139	3	US-10-799-870-22	Sequence 22, Appl
C 32	19.2	76.8	25992	3	US-09-949-016-17308	Sequence 17308, A
C 33	19.2	76.8	45319	3	US-09-949-016-12526	Sequence 12526, A
C 34	19.2	76.8	46323	3	US-09-949-016-13267	Sequence 13267, A
C 35	19.2	76.8	55298	3	US-09-491-356C-1	Sequence 1, Appli
C 36	19.2	76.8	88490	3	US-09-949-016-12758	Sequence 12758, A
C 37	19.2	76.8	88736	3	US-09-949-016-14222	Sequence 14222, A
C 38	19.2	76.8	157866	3	US-09-949-016-12982	Sequence 12982, A
C 39	19.2	76.8	157866	3	US-09-949-016-12983	Sequence 12983, A
C 40	19.2	76.8	157866	3	US-09-949-016-12984	Sequence 12984, A
C 41	19.2	76.8	187169	3	US-09-949-016-12776	Sequence 12776, A
C 42	19.2	76.8	191569	3	US-09-949-016-15940	Sequence 15940, A
C 43	19.2	76.8	390890	3	US-09-949-016-14720	Sequence 14720, A
C 44	18.8	75.2	601	3	US-09-949-016-153451	Sequence 153451, A
C 45	18.8	75.2	601	3	US-09-949-016-153452	Sequence 153452, A
C 46	18.8	75.2	601	3	US-09-949-016-204628	Sequence 204628, A
C 47	18.8	75.2	1227	3	US-09-248-796A-6555	Sequence 6555, Ap
C 48	18.8	75.2	1336	3	US-09-257-179-13	Sequence 13, Appl
C 49	18.8	75.2	2339	3	US-09-270-767-24171	Sequence 24171, A
C 50	18.8	75.2	2339	3	US-09-270-767-24171	Sequence 24171, A
C 51	18.8	75.2	122626	3	US-09-949-016-17524	Sequence 17524, A
C 52	18.8	75.2	130563	3	US-09-949-016-12273	Sequence 12273, A
C 53	18.8	75.2	131379	3	US-09-949-016-16050	Sequence 16050, A
C 54	18.6	74.4	27	2	US-08-484-816-17	Sequence 17, Appl
C 55	18.6	74.4	27	2	US-08-476-625-17	Sequence 17, Appl
C 56	18.6	74.4	27	2	US-08-949-076-17	Sequence 17, Appl
C 57	18.6	74.4	27	2	US-08-484-519-17	Sequence 17, Appl
C 58	18.6	74.4	55	2	US-08-484-816-13	Sequence 13, Appl
C 59	18.6	74.4	55	2	US-08-484-816-15	Sequence 15, Appl
C 60	18.6	74.4	55	2	US-08-476-625-13	Sequence 13, Appl
C 61	18.6	74.4	55	2	US-08-476-625-15	Sequence 15, Appl
C 62	18.6	74.4	55	2	US-08-949-076-13	Sequence 13, Appl
C 63	18.6	74.4	55	2	US-08-949-076-15	Sequence 15, Appl
C 64	18.6	74.4	55	2	US-08-484-519-13	Sequence 13, Appl
C 65	18.6	74.4	55	2	US-08-484-519-15	Sequence 15, Appl
C 66	18.6	74.4	63	3	US-09-513-999C-26842	Sequence 26842, A
C 67	18.6	74.4	83	3	US-09-270-767-30460	Sequence 30460, A
C 68	18.6	74.4	95	2	US-08-484-816-14	Sequence 14, Appl
C 69	18.6	74.4	95	2	US-08-476-625-14	Sequence 14, Appl
C 70	18.6	74.4	95	2	US-08-949-076-14	Sequence 14, Appl
C 71	18.6	74.4	95	2	US-08-484-519-14	Sequence 14, Appl
C 72	18.6	74.4	172	3	US-09-513-999C-20783	Sequence 20783, A
C 73	18.6	74.4	232	3	US-09-621-976-10815	Sequence 10815, A
C 74	18.6	74.4	601	3	US-09-949-016-24151	Sequence 24151, A
C 75	18.6	74.4	601	3	US-09-949-016-23713	Sequence 23713, A
C 76	18.6	74.4	601	3	US-09-949-016-73714	Sequence 73714, A
C 77	18.6	74.4	601	3	US-09-949-016-10786	Sequence 10786, A
C 78	18.6	74.4	601	3	US-09-949-016-112503	Sequence 112503, A
C 79	18.6	74.4	601	3	US-09-949-016-124192	Sequence 124192, A
C 80	18.6	74.4	601	3	US-09-949-016-124193	Sequence 124193, A
C 81	18.6	74.4	601	3	US-09-949-016-158191	Sequence 158191, A
C 82	18.6	74.4	601	3	US-09-949-016-160712	Sequence 160712, A
C 83	18.6	74.4	601	3	US-09-949-016-160713	Sequence 160713, A
C 84	18.6	74.4	601	3	US-09-949-016-160714	Sequence 160714, A
C 85	18.6	74.4	611	3	US-09-404-879A-343	Sequence 343, App
C 86	18.6	74.4	611	3	US-09-667-857-343	Sequence 343, App
C 87	18.6	74.4	611	3	US-10-196-053-343	Sequence 343, App
C 88	18.6	74.4	611	3	US-09-827-271-343	Sequence 343, App
C 89	18.6	74.4	665	3	US-09-270-767-14317	Sequence 14317, A
C 90	18.6	74.4	2685	3	US-10-104-047-1143	Sequence 1143, Ap
C 91	18.6	74.4	2699	3	US-09-949-016-15525	Sequence 15525, Ap
C 92	18.6	74.4	2704	3	US-09-949-016-784	Sequence 784, App
C 93	18.6	74.4	3073	2	US-07-688-352C-31	Sequence 31, Appl
C 94	18.6	74.4	3073	2	US-08-474-379C-31	Sequence 31, Appl
C 95	18.6	74.4	3073	2	US-09-146-249A-31	Sequence 31, Appl
C 96	18.6	74.4	3073	3	US-08-206-188B-31	Sequence 31, Appl
C 97	18.6	74.4	3073	6	PCT-US91-02714-30	Sequence 30, Appl

98 18.6 74.4 3479 3 US-08-714-918-7
99 18.6 74.4 3479 3 US-09-265-315-7
100 18.6 74.4 3479 3 US-09-265-315-7
101 18.6 74.4 3479 3 US-09-268-417-7
102 18.6 74.4 3479 3 US-09-528-709-7
103 18.6 74.4 3479 3 US-09-527-745-7
104 18.6 74.4 4881 3 US-09-949-016-13299
105 18.6 74.4 5924 3 US-08-956-171E-130
106 18.6 74.4 5924 3 US-08-781-986A-130
107 18.6 74.4 5925 3 US-08-470-260-4
108 18.6 74.4 5925 3 US-08-471-491-4
109 18.6 74.4 5925 3 US-08-466-662-4
110 18.6 74.4 9772 3 US-09-949-016-17211
111 18.6 74.4 10299 2 US-08-477-451-1
112 18.6 74.4 10299 2 US-08-477-451-1
113 18.6 74.4 15216 3 US-09-949-016-17377
114 18.6 74.4 17056 3 US-09-245-041-3
115 18.6 74.4 17056 3 US-09-358-055B-3
116 18.6 74.4 17056 3 US-09-893-238-3
117 18.6 74.4 19932 2 US-08-477-451-25
118 18.6 74.4 23383 3 US-09-949-016-13726
119 18.6 74.4 23394 3 US-09-949-016-13887
120 18.6 74.4 32392 3 US-09-662-254B-27
121 18.6 74.4 35064 3 US-09-949-016-12778
122 18.6 74.4 35065 3 US-09-949-016-13196
123 18.6 74.4 36577 3 US-09-949-016-13176
124 18.6 74.4 48994 3 US-09-949-016-14091
125 18.6 74.4 50836 3 US-09-949-016-16722
126 18.6 74.4 52636 3 US-09-949-016-14839
127 18.6 74.4 54878 3 US-09-949-016-12255
128 18.6 74.4 59479 3 US-09-949-016-16910
129 18.6 74.4 60304 3 US-09-949-016-12218
130 18.6 74.4 60305 3 US-09-949-016-15791
131 18.6 74.4 63187 3 US-09-949-016-12682
132 18.6 74.4 63187 3 US-09-949-016-16288
133 18.6 74.4 71251 3 US-09-949-016-15332
134 18.6 74.4 77036 3 US-09-949-016-16156
135 18.6 74.4 82125 3 US-09-949-016-13517
136 18.6 74.4 82125 3 US-09-949-016-13518
137 18.6 74.4 83349 3 US-09-949-016-17149
138 18.6 74.4 84132 3 US-09-949-016-16241
139 18.6 74.4 88557 3 US-09-949-016-17028
140 18.6 74.4 94095 3 US-09-949-016-14389
141 18.6 74.4 100928 3 US-09-949-016-16926
142 18.6 74.4 101558 3 US-09-949-016-12243
143 18.6 74.4 113966 3 US-09-949-016-12277
144 18.6 74.4 113967 3 US-09-949-016-17051
145 18.6 74.4 120213 3 US-09-949-016-13304
146 18.6 74.4 120217 3 US-09-949-016-12260
147 18.6 74.4 143644 3 US-09-949-016-15238
148 18.6 74.4 168104 3 US-09-949-016-12026
149 18.6 74.4 168105 3 US-09-949-016-16554
150 18.6 74.4 174170 3 US-09-949-016-14810

ALIGNMENTS

Sequence 7, Appl1
Sequence 7, Appl1
Sequence 7, Appl1
Sequence 7, Appl1
Sequence 7, Appl1
Sequence 7, Appl1
Sequence 13299, A
Sequence 130, App
Sequence 130, App
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 1711, A
Sequence 1, Appl1
Sequence 5, Appl1
Sequence 17377, A
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 25, Appl1
Sequence 13726, A
Sequence 13887, A
Sequence 27, Appl1
Sequence 13196, A
Sequence 12778, A
Sequence 13176, A
Sequence 14091, A
Sequence 16722, A
Sequence 14839, A
Sequence 12255, A
Sequence 16910, A
Sequence 12218, A
Sequence 15791, A
Sequence 16282, A
Sequence 16288, A
Sequence 15332, A
Sequence 16156, A
Sequence 13517, A
Sequence 13518, A
Sequence 17149, A
Sequence 16241, A
Sequence 17028, A
Sequence 14389, A
Sequence 16926, A
Sequence 12243, A
Sequence 12277, A
Sequence 17051, A
Sequence 13304, A
Sequence 12260, A
Sequence 15238, A
Sequence 12026, A
Sequence 16554, A
Sequence 14810, A

RESULT 1
US-09-949-016-15095/c
; Sequence 15095, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15095
; LENGTH: 93510
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(93510)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15095

Query Match 83.2%; Score 20.8; DB 3; Length 93510;
Best Local Similarity 91.7%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTGATCT 24
Db 18619 AAAAAAAAACTAAAGCTGATCT 18596

RESULT 2
US-09-270-767-9667/c
; Sequence 9667, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1998-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9667
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9667

Query Match 80.8%; Score 20.2; DB 3; Length 627;
Best Local Similarity 88.0%; Pred. No. 1,4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTGATCT 25
Db 122 AAAAAAAAACTAAAGCTGATATT 98

RESULT 3
US-09-270-767-24949/c
; Sequence 24949, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24949
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24949

Query Match 80.8%; Score 20.2; DB 3; Length 627;
Best Local Similarity 88.0%; Pred. No. 1,4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTGATCT 25

Db 122 AAAACAAAACCTAAGCTGATATT 98

RESULT 4

US-09-543-681A-4161/C
; Sequence 4161, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4161
; LENGTH: 945
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-4161

Query Match 79.2%; Score 19.8; DB 3; Length 945;
Best Local Similarity 91.3%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAATACTAAGCTTGATC 23

Db 653 AAACAAAACCTAAGCTTGATC 631

RESULT 5

US-09-270-767-9472/C
; Sequence 9472, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9472
; LENGTH: 1184
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9472

Query Match 79.2%; Score 19.8; DB 3; Length 1184;
Best Local Similarity 91.3%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAATACTAAGCTTGATC 23

Db 61 AAAAATACTAAGCTTGATC 39

RESULT 6

US-09-270-767-24754/C
; Sequence 24754, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24754

; LENGTH: 1184
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24754

Query Match 79.2%; Score 19.8; DB 3; Length 1184;
Best Local Similarity 91.3%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAATACTAAGCTTGATC 23

Db 61 AAAAATACTAAGCTTGATC 39

RESULT 7

US-08-851-843A-102/C
; Sequence 102, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Hatley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809e1 Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,843A
; FILING DATE: 06-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-851-843A-102

Query Match 76.8%; Score 19.2; DB 3; Length 50;

Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCT 24
Db 50 AAAAAAAAAAAGCTTGACT 27

RESULT 8

US-08-974-549A-587/C
Sequence 587, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015399-002610US
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 587:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..50
OTHER INFORMATION: /note="Q-T primer"

US-08-974-549A-587

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCT 24
Db 50 AAAAAAAAAAAGCTTGACT 27

RESULT 9

US-08-854-050-102/C
Sequence 102, Application US/08854050
Patent No. 6261836
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-854-050-102

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
|||
Db 50 AAAAAAAAAAAAGCTTGAGCT 27

RESULT 10
US-09-430-323-102/C
Sequence 102, Application US/09430323
Patent No. 6309867
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Langner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin H.
Andrews, William H.
TITLE OF INVENTION: No. 6309867el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,323
FILING DATE: 29-Oct-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-430-323-102

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
|||
Db 50 AAAAAAAAAAAAGCTTGAGCT 27

RESULT 11
US-09-402-181B-587/C
Sequence 587, Application US/09402181B
Patent No. 6610839
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Langner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin H.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 633
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/402,181B
FILING DATE: 29-Sep-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ansehnus, Scott L.
REGISTRATION NUMBER: 42,271
REFERENCE/DOCKET NUMBER: 015389-002620US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 587:

SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..50
OTHER INFORMATION: /note="Q-T primer"
SEQUENCE DESCRIPTION: SEQ ID NO: 587:
US-09-402-181B-587

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCT 24
Db 50 AAAAAAAAAAAGCTTGAGCT 27

RESULT 12
US-09-721-456-587/c
Sequence 587, Application US/09721456
Patent No. 6617110
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/721,456
FILING DATE: 22-Nov. 6617110-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-Nov-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-Oct-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-Apr-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-Apr-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-May-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-May-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-Aug-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-Aug-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-Aug-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-Oct-1997

APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-Oct-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 587:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..50
OTHER INFORMATION: /note="Q-T primer"
SEQUENCE DESCRIPTION: SEQ ID NO: 587:
US-09-721-456-587

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCT 24
Db 50 AAAAAAAAAAAGCTTGAGCT 27

RESULT 13
US-09-766-253-102/c
Sequence 102, Application US/09766253
Patent No. 6808880
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin
Andrews, William H.
TITLE OF INVENTION: No. 6808880el Telomerase
NUMBER OF SEQUENCES: 171
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/766,253
FILING DATE: 19-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/846,017
FILING DATE: 1997-04-25
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-Oct-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002920US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-766-253-102

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
|||||
Db 50 AAAAAAAAAAAGCTTGAGCT 27

RESULT 14
US-10-054-295-102/c
Sequence 102, Application US/10054295
Patent No. 6921664
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin
Andrews, William H.
TITLE OF INVENTION: No. 6921664e1 Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/054,295
FILING DATE: 18-Jan-2002
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/854,050
FILING DATE: <unknown>
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-10-054-295-102

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
|||||
Db 50 AAAAAAAAAAAGCTTGAGCT 27

RESULT 15
US-09-438-486A-102/c
Sequence 102, Application US/09438486A
Patent No. 6927285
GENERAL INFORMATION:
APPLICANT: CECH, THOMAS R.
LINGNER, JOACHIM
NAKAMURA, TORU
APPLICANT: CHAPMAN, KAREN B.
MORIN, GREGG B.
APPLICANT: HARLEY, CALVIN
APPLICANT: ANDREWS, WILLIAM H.
TITLE OF INVENTION: GENE FOR HUMAN TELOMERASE REVERSE TRANSCRIPTASE AND
FILE REFERENCE: 018/062
CURRENT APPLICATION NUMBER: US/09/438,486A
CURRENT FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 08/851,843
PRIOR FILING DATE: 1997-05-06
PRIOR APPLICATION NUMBER: 08/846,017
PRIOR FILING DATE: 1997-04-25
PRIOR APPLICATION NUMBER: 08/844,419
PRIOR FILING DATE: 1997-04-18
PRIOR APPLICATION NUMBER: 08/724,643
PRIOR FILING DATE: 1996-10-01
NUMBER OF SEQ ID NOS: 223
SOFTWARE: Patentin Ver. 3.2
SEQ ID NO 102
LENGTH: 50
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide
US-09-438-486A-102

Query Match 76.8%; Score 19.2; DB 3; Length 50;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
|||||
Db 50 AAAAAAAAAAAGCTTGAGCT 27

RESULT 16
US-09-057-328-1/c
Sequence 1, Application US/09057328
Patent No. 6080576
GENERAL INFORMATION:
APPLICANT: Zambrowicz, Brian
APPLICANT: Friedlich, Glenn
APPLICANT: Sands, Arthur T.
TITLE OF INVENTION: VECTORS FOR GENE TRAPPING
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA

```
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/057,328
FILING DATE: 08-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cornuzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 8535-0020-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 52 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-057-328-1
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
    |||||
DB 50 AAAAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 17
US-09-258-373-11/C
; Sequence 11, Application US/09258373
; Patent No. 6150110
; GENERAL INFORMATION:
; APPLICANT: Fletcher, Jonathan A.
; APPLICANT: Xiao, Sheng
; TITLE OF INVENTION: HMG(Y)-LMA4+ FUSION ONCOGENE,
; FILE REFERENCE: B0801/7135/ERP
; CURRENT APPLICATION NUMBER: US/09/258,373
; EARLIER FILING DATE: 1999-02-26
; EARLIER APPLICATION NUMBER: 60/076,401
; EARLIER FILING DATE: 1998-02-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-258-373-11
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
    |||||
DB 50 AAAAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 18
US-09-091-590A-15/C
; Sequence 15, Application US/09091590A
; Patent No. 6242574
```

```
GENERAL INFORMATION:
APPLICANT: Nielsen, Klaus
APPLICANT: Kroil Kristensen, Anne
APPLICANT: Brunstedt, Janne
TITLE OF INVENTION: Anti-Microbial Proteins
FILE REFERENCE: S-137-1101/MA/A/SGS/PCT
CURRENT APPLICATION NUMBER: US/09/091,590A
CURRENT FILING DATE: 1999-05-06
PRIOR APPLICATION NUMBER: PCT/EP96/05765
PRIOR FILING DATE: 1996-12-20
PRIOR APPLICATION NUMBER: GB 9526238.2
PRIOR FILING DATE: 1995-12-21
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn version 3.0
SEQ ID NO 15
LENGTH: 52
TYPE: DNA
ORGANISM: Artificial/Unknown
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(52)
OTHER INFORMATION: primer
US-09-091-590A-15
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
    |||||
DB 50 AAAAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 19
US-09-228-455-1/C
; Sequence 1, Application US/09228455
; Patent No. 6413720
; GENERAL INFORMATION:
; APPLICANT: Pardini, Jose R.
; APPLICANT: Chan, Kyle W.H.
; TITLE OF INVENTION: METHOD FOR THE AFFINITY ISOLATION OF NEWLY SYNTHESIZED
; FILE REFERENCE: 860098,428
; CURRENT APPLICATION NUMBER: US/09/228,455
; EARLIER FILING DATE: 1999-01-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
US-09-228-455-1
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
    |||||
DB 50 AAAAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 20
US-09-228-455-2
; Sequence 2, Application US/09228455
; Patent No. 6413720
; GENERAL INFORMATION:
; APPLICANT: Pardini, Jose R.
; APPLICANT: Chan, Kyle W.H.
; TITLE OF INVENTION: METHOD FOR THE AFFINITY ISOLATION OF NEWLY SYNTHESIZED
```

```
; TITLE OF INVENTION: RNA
; FILE REFERENCE: 860098.428
; CURRENT APPLICATION NUMBER: US/09/228.455
; CURRENT FILING DATE: 1999-01-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
US-09-228-455-2
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 3 AAAAAAAAAAAAAAAAAAGCTTGAGCT 26
```

```
RESULT 21
US-09-276-533A-17/c
; Sequence 17, Application US/09276533A
; Patent No. 6436707
; GENERAL INFORMATION:
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Friedlich, Glenn A.
; APPLICANT: Lilleberg, Stan
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: VECTORS FOR GENE MUTAGENESIS AND GENE
; FILE REFERENCE: 07705-0006-00000
; CURRENT APPLICATION NUMBER: US/09/276, 533A
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: US 60/079, 729
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/081, 727
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: US 60/109, 302
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-09-276-533A-17
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 50 AAAAAAAAAAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 22
US-10-158-735-17/c
; Sequence 17, Application US/10158735
; Patent No. 6776988
; GENERAL INFORMATION:
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Friedlich, Glenn A.
; APPLICANT: Lilleberg, Stan
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: VECTORS FOR GENE MUTAGENESIS AND GENE
```

```
; TITLE OF INVENTION: DISCOVERY
; FILE REFERENCE: 7705-0006-01
; CURRENT APPLICATION NUMBER: US/10/158, 735
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 09/276, 533
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/079, 729
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/081, 727
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: US 60/109, 302
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-10-158-735-17
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 50 AAAAAAAAAAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 23
US-09-443-282B-12/c
; Sequence 12, Application US/09443282B
; Patent No. 6808921
; GENERAL INFORMATION:
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Friedlich, Glenn A.
; APPLICANT: Lilleberg, Stan
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Vectors for Gene Mutagenesis and Gene Discovery
; FILE REFERENCE: 07705-0001-00000
; CURRENT APPLICATION NUMBER: US/09/443, 282B
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 09/276, 533
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/109, 302
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: US 60/081, 727
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: US 60/079, 729
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 52
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric sequence
US-09-443-282B-12
```

```
Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 50 AAAAAAAAAAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 24
US-10-154-517-1/c
```

```
/ Sequence 1, Application US/10154517
/ Patent No. 6960438
/ GENERAL INFORMATION:
/ APPLICANT: Pardini, Jose R.
/ APPLICANT: Chan, Kyle W.H.
/ TITLE OF INVENTION: METHOD FOR THE AFFINITY ISOLATION OF NEWLY SYNTHESIZED
/ TITLE OF INVENTION: RNA
/ FILE REFERENCE: 860098.428
/ CURRENT APPLICATION NUMBER: US/10/154,517
/ CURRENT FILING DATE: 2002-05-23
/ PRIOR APPLICATION NUMBER: US/09/228,455
/ PRIOR FILING DATE: 1999-01-11
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: Patent Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 52
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:
/ OTHER INFORMATION: Deoxyoligonucleotide primer.
US-10-154-517-1

Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGAATCT 24
Db 50 AAAAAAAAAAAGCTTGAATCT 27

RESULT 25
US-10-154-517-2
/ Sequence 2, Application US/10154517
/ Patent No. 6960438
/ GENERAL INFORMATION:
/ APPLICANT: Pardini, Jose R.
/ APPLICANT: Chan, Kyle W.H.
/ TITLE OF INVENTION: METHOD FOR THE AFFINITY ISOLATION OF NEWLY SYNTHESIZED
/ TITLE OF INVENTION: RNA
/ FILE REFERENCE: 860098.428
/ CURRENT APPLICATION NUMBER: US/10/154,517
/ CURRENT FILING DATE: 2002-05-23
/ PRIOR APPLICATION NUMBER: US/09/228,455
/ PRIOR FILING DATE: 1999-01-11
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: Patent Ver. 2.0
/ SEQ ID NO 2
/ LENGTH: 52
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:
/ OTHER INFORMATION: Deoxyoligonucleotide primer.
US-10-154-517-2

Query Match          76.8%; Score 19.2; DB 3; Length 52;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/ FILE REFERENCE: GENSET.054PR2
/ CURRENT APPLICATION NUMBER: US/09/621,976
/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 17988
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-621-976-17988

Query Match          76.8%; Score 19.2; DB 3; Length 304;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGAATCT 24
Db 300 AAAAAAAAAAAGCTTGAATCT 277

RESULT 27
US-09-949-016-52788/C
/ Sequence 52788, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 52788
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-52788

Query Match          76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGAATCT 24
Db 295 AAAAAAAAAAAGCTTGAATCT 272

RESULT 28
US-09-328-111-342/C
/ Sequence 342, Application US/09328111
/ Patent No. 6262333
/ GENERAL INFORMATION:
/ APPLICANT: Endege, Wilson O.
/ APPLICANT: Steinmann, Kathleen E.
/ APPLICANT: Aetle, Jon H.
/ APPLICANT: Burgess, Christopher C.
/ APPLICANT: Bushnell, Steven E.
/ APPLICANT: Carroll III, Eddie
/ APPLICANT: Catino, Theodore J.
/ APPLICANT: Detli, Adnan
/ APPLICANT: Ford, Donna M.
/ APPLICANT: Lewis, Marcia E.
/ APPLICANT: Monahan, John E.
/ APPLICANT: Schlegel, Robert
```

```
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 342
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(669)
; OTHER INFORMATION: n = A,T,C or G
; US-09-328-111-342

Query Match          76.8%; Score 19.2; DB 3; Length 669;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 26 AAAAAAAAAAAAAAAAAAGCTTGACCT 3

RESULT 29
US-09-322-478-22
; Sequence 22, Application US/09322478
; Patent No. 6331662
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
; FILE REFERENCE: P-1065 ISURF Plant Retroelement
; CURRENT APPLICATION NUMBER: US/09/322,478
; CURRENT FILING DATE: 1999-05-28
; EARLIER APPLICATION NUMBER: 60/087125
; EARLIER FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 9139
; TYPE: DNA
; ORGANISM: Glycine max
; US-09-322-478-22

Query Match          76.8%; Score 19.2; DB 3; Length 9139;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 7244 AAAAAAAAAAGTAAAGTTATCT 7267

RESULT 30
US-09-586-106D-22
; Sequence 22, Application US/09586106D
; Patent No. 6720479
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: PLANT RETROELEMENTS AND METHODS RELATED THERETO
; FILE REFERENCE: P-1065A
; CURRENT APPLICATION NUMBER: US/09/586,106D
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/087,125
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 09/322,478
; PRIOR FILING DATE: 1999-05-28
```

```
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 9139
; TYPE: DNA
; ORGANISM: Glycine max
; US-09-586-106D-22

Query Match          76.8%; Score 19.2; DB 3; Length 9139;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 7244 AAAAAAAAAAGTAAAGTTATCT 7267

RESULT 31
US-10-799-870-22
; Sequence 22, Application US/10799870
; Patent No. 6949695
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: PLANT RETROELEMENTS AND METHODS RELATED THERETO
; FILE REFERENCE: P-1065A
; CURRENT APPLICATION NUMBER: US/10/799,870
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US/09/586,106
; PRIOR FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/087,125
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 09/322,478
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 9139
; TYPE: DNA
; ORGANISM: Glycine max
; US-10-799-870-22

Query Match          76.8%; Score 19.2; DB 3; Length 9139;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 7244 AAAAAAAAAAGTAAAGTTATCT 7267

RESULT 32
US-09-949-016-17308/C
; Sequence 22, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17308
; LENGTH: 25992
; TYPE: DNA
```

ORGANISM: Human
US-09-949-016-17308

Query Match 76.8%; Score 19.2; DB 3; Length 25992;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTGATCT 24
Db 25314 AAAAAAAAACTAAGCTGATCT 25291

RESULT 33
US-09-949-016-12526/C
Sequence 12526, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12526
LENGTH: 46319
TYPE: DNA
ORGANISM: Human
US-09-949-016-12526

Query Match 76.8%; Score 19.2; DB 3; Length 46319;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTGATCT 24
Db 7170 AAAAAAAAACTAAGCTGATCT 7147

RESULT 34
US-09-949-016-13267/C
Sequence 13267, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13267
LENGTH: 46323
TYPE: DNA
ORGANISM: Human
US-09-949-016-13267

Query Match 76.8%; Score 19.2; DB 3; Length 46323;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTGATCT 24
Db 7170 AAAAAAAAACTAAGCTGATCT 7147

RESULT 35
US-09-491-356C-1
Sequence 1, Application US/09491356C
Patent No. 6566061
GENERAL INFORMATION:
APPLICANT: Philibert, Robert A.
APPLICANT: Gims, Edward I.
TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF X013
FILE REFERENCE: 9465.6US11
CURRENT APPLICATION NUMBER: US/09/491,356C
CURRENT FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: PCT/US99/09365
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,465
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 55298
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (485)..(485)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (838)..(838)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (16728)..(16728)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (22750)..(22750)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (22756)..(22756)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (28519)..(28519)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (44804)..(44804)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (45002)..(45002)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (54049)..(54049)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (54226)..(54226)
OTHER INFORMATION: n is not determined
US-09-491-356C-1

Query Match 76.8%; Score 19.2; DB 3; Length 55298;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTGATCT 24
Db 16884 AAAAAAAAACTAAGCTGATCT 16907

RESULT 36
US-09-949-016-12758/C


```
; Sequence 12758, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12758
; LENGTH: 88490
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12758

Query Match          76.8%; Score 19.2; DB 3; Length 88490;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 35378 AAAAAAAAACTAAAGCATTAAC 35355

RESULT 37
US-09-949-016-14222/c
; Sequence 14222, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14222
; LENGTH: 88736
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14222

Query Match          76.8%; Score 19.2; DB 3; Length 88736;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 35624 AAAAAAAAACTAAAGCATTAAC 35601

RESULT 38
US-09-949-016-12982
; Sequence 12982, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12982
; LENGTH: 157866
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12982

Query Match          76.8%; Score 19.2; DB 3; Length 157866;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 86898 AAAAAAAAAAGTAAAGCTTGATCT 86921

RESULT 39
US-09-949-016-12983
; Sequence 12983, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12983
; LENGTH: 157866
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12983

Query Match          76.8%; Score 19.2; DB 3; Length 157866;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 86898 AAAAAAAAAAGTAAAGCTTGATCT 86921

RESULT 40
US-09-949-016-12984
; Sequence 12984, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 12984
/ LENGTH: 157866
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-12984

Query Match          76.8%; Score 19.2; DB 3; Length 157866;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTGATCT 24
Db 86898 AAAAAAAAAAGTAAAGCTGATCT 86921

RESULT 41
US-09-949-016-12776/C
/ Sequence 12776, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 12776
/ LENGTH: 187169
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(187169)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

Query Match          76.8%; Score 19.2; DB 3; Length 187169;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAAGCTGATCT 25
Db 42190 AAAAAAAAAATCTAAATGATCT 42167

RESULT 42
US-09-949-016-15940/C
/ Sequence 15940, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
```

```
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 15940
/ LENGTH: 191569
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(191569)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

Query Match          76.8%; Score 19.2; DB 3; Length 191569;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAAGCTGATCT 25
Db 42190 AAAAAAAAAATCTAAATGATCT 42167

RESULT 43
US-09-949-016-14720/C
/ Sequence 14720, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 14720
/ LENGTH: 390890
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(390890)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14720

Query Match          76.8%; Score 19.2; DB 3; Length 390890;
Best Local Similarity 87.5%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTGATCT 24
Db 103002 AAAAAAAAAATCTAAATGATCT 102979

RESULT 44
US-09-949-016-153451
/ Sequence 153451, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153451
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-153451
```

```
Query Match      75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 AAAAAAAAACTAAAGCTTGAT 22
Db      287 AAAAAAAAAAAAAAAAAAGCTTGAT 308
```

```
RESULT 45
US-09-949-016-153452
; Sequence 153452; Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153452
; LENGTH: 601.
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-153452
```

```
Query Match      75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 AAAAAAAAACTAAAGCTTGAT 22
Db      505 AAAAAAAAAAAAAAAAAAGCTTGAT 526
```

```
RESULT 46
US-09-949-016-204628/c
; Sequence 204628; Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 204628
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-204628
```

```
Query Match      75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 AAAAAAAAACTAAAGCTTGAT 22
Db      553 AAAAAAAAAATTAAGCTTGAT 532
```

```
RESULT 47
US-09-248-796A-6555/c
; Sequence 6555; Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICP
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 6555
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-6555
```

```
Query Match      75.2%; Score 18.8; DB 3; Length 1227;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      4 AAAAAAAAACTAAAGCTTGATCTT 25
Db      632 AAAAAAAAACTAATGTTGATCTT 611
```

```
RESULT 48
US-09-257-179-13/c
; Sequence 13; Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: P2015P1
; CURRENT APPLICATION NUMBER: US/09/257,179
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: PCT/US98/17709
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: 60/056,270
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/056,271
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/056,247
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/056,073
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
```

Job time : 60.5 secs

```

; LENGTH: 1336
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (766)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-257-179-13
```

```
Query Match          75.2%; Score 18.8; DB 3; Length 1336;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAAGCTTGAT 22
   |||||
Db 594 AAAAAAAAACTAAAGCTTGAT 573
```

```
RESULT 49
US-09-270-767-9189
; Sequence 9189, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 9189
; LENGTH: 2339
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9189
```

```
Query Match          75.2%; Score 18.8; DB 3; Length 2339;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAAGCTTGAT 22
   |||||
Db 702 AAAAAAAAACTAAATCTTGT 723
```

```
RESULT 50
US-09-270-767-24471
; Sequence 24471, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 24471
; LENGTH: 2339
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24471
```

```
Query Match          75.2%; Score 18.8; DB 3; Length 2339;
Best Local Similarity 90.9%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAAGCTTGAT 22
   |||||
Db 702 AAAAAAAAACTAAATCTTGT 723
```

Search completed: December 14, 2005, 07:44:20

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds

(without alignments)
800.703 Million cell updates/sec

Title: US-10-681-773-4

Perfect score: 25
Sequence: 1 aaaaaaaaaactagctgacctc 25

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapect 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/1_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/H_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PCOMB_COMB.seq: *
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq: *
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq: *
9: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.8	79.2	601	3	US-09-949-016-33868
2	19.8	79.2	601	3	US-09-949-016-96073
3	19.8	79.2	387902	3	US-09-949-016-14543
4	19.8	79.2	421883	3	US-09-949-016-12557
5	19.2	76.8	377	3	US-09-270-767-5200
6	19.2	76.8	377	3	US-09-270-767-20482
7	19.2	76.8	93510	3	US-09-949-016-15095
8	19.2	76.8	1117	3	US-08-858-2076-40
9	18.8	75.2	2727	3	US-09-248-796A-638
10	18.8	75.2	18264	3	US-09-949-016-16469
11	18.8	75.2	74790	3	US-09-949-016-15311
12	18.6	74.4	601	3	US-09-949-016-20819
13	18.6	74.4	601	3	US-09-949-016-67144
14	18.6	74.4	601	3	US-09-949-016-85277
15	18.6	74.4	601	3	US-09-949-016-86539
16	18.6	74.4	627	3	US-09-270-767-9667
17	18.6	74.4	627	3	US-09-270-767-24949
18	18.6	74.4	8839	3	US-09-949-016-14171
19	18.6	74.4	12717	3	US-09-693-205A-9
20	18.6	74.4	12717	3	US-09-693-205A-13
21	18.6	74.4	12717	3	US-09-693-205A-14
22	18.6	74.4	12792	3	US-09-693-205A-7
23	18.6	74.4	12793	3	US-09-693-205A-11
24	18.6	74.4	12793	3	US-09-693-205A-11

C 25	18.6	74.4	12793	3	US-09-693-205A-12	Sequence 12, App1
C 26	18.6	74.4	12793	3	US-09-693-205A-15	Sequence 15, App1
C 27	18.6	74.4	13115	3	US-09-949-016-11895	Sequence 11895, A
C 28	18.6	74.4	13371	3	US-09-949-016-13723	Sequence 13723, A
C 29	18.6	74.4	14000	3	US-09-817-762-10	Sequence 10, App1
C 30	18.6	74.4	50553	3	US-09-949-016-15821	Sequence 15821, A
C 31	18.6	74.4	51849	3	US-09-949-016-15433	Sequence 15433, A
C 32	18.6	74.4	93920	3	US-09-949-016-12461	Sequence 12461, A
C 33	18.6	74.4	93920	3	US-09-949-016-16853	Sequence 16853, A
C 34	18.6	74.4	183112	3	US-09-949-016-14184	Sequence 14184, A
C 35	18.6	74.4	187595	3	US-09-949-016-15546	Sequence 15546, A
C 36	18.6	74.4	283538	3	US-09-949-016-13506	Sequence 13506, A
C 37	18.6	74.4	314798	3	US-09-949-016-11359	Sequence 11359, A
C 38	18.4	73.6	105919	3	US-09-949-016-11769	Sequence 11769, A
C 39	18.2	72.8	246	3	US-09-248-796A-10034	Sequence 10034, A
C 40	18.2	72.8	601	3	US-09-949-016-19777	Sequence 19777, A
C 41	18.2	72.8	601	3	US-09-949-016-20890	Sequence 20890, A
C 42	18.2	72.8	945	3	US-09-543-681A-4161	Sequence 4161, Ap
C 43	18.2	72.8	1184	3	US-09-270-767-9472	Sequence 9472, Ap
C 44	18.2	72.8	1184	3	US-09-270-767-24754	Sequence 24754, A
C 45	18.2	72.8	63467	3	US-09-949-002-693	Sequence 693, App
C 46	18.2	72.8	64518	3	US-09-949-016-17289	Sequence 17289, A
C 47	18.2	72.8	64518	3	US-09-949-016-15530	Sequence 15530, A
C 48	18.2	72.8	112705	3	US-09-248-796A-13477	Sequence 13477, A
C 49	17.8	71.2	234	3	US-09-621-976-17000	Sequence 900, App
C 50	17.8	71.2	1008	3	US-09-134-000C-900	Sequence 28, App1
C 51	17.8	71.2	1227	3	US-09-830-230A-28	Sequence 22, App1
C 52	17.8	71.2	1333	3	US-09-830-230A-27	Sequence 22, App1
C 53	17.8	71.2	1539	2	US-08-899-811-22	Sequence 391, App
C 54	17.8	71.2	1773	3	US-09-248-796A-331	Sequence 16211, A
C 55	17.8	71.2	1773	3	US-09-949-016-16212	Sequence 16211, A
C 56	17.8	71.2	16056	3	US-09-949-016-16212	Sequence 15680, A
C 57	17.8	71.2	16056	3	US-09-949-016-15680	Sequence 15680, A
C 58	17.8	71.2	46725	3	US-09-949-016-12530	Sequence 12530, A
C 59	17.8	71.2	64291	3	US-09-949-016-17321	Sequence 17321, A
C 60	17.8	71.2	109250	3	US-09-949-016-14624	Sequence 14624, A
C 61	17.8	71.2	109251	3	US-09-949-016-16274	Sequence 16274, A
C 62	17.8	71.2	117410	3	US-08-851-843A-102	Sequence 102, App
C 63	17.8	71.2	147835	3	US-08-854-050-102	Sequence 102, App
C 64	17.8	70.4	50	3	US-09-430-323-102	Sequence 587, App
C 65	17.6	70.4	50	3	US-09-402-181B-587	Sequence 587, App
C 66	17.6	70.4	50	3	US-09-721-456-587	Sequence 587, App
C 67	17.6	70.4	50	3	US-09-765-253-102	Sequence 102, App
C 68	17.6	70.4	50	3	US-10-054-295-102	Sequence 102, App
C 69	17.6	70.4	50	3	US-09-438-486A-102	Sequence 102, App
C 70	17.6	70.4	50	3	US-09-057-328-11	Sequence 11, App1
C 71	17.6	70.4	50	3	US-09-057-328-11	Sequence 11, App1
C 72	17.6	70.4	50	3	US-09-091-590A-15	Sequence 15, App1
C 73	17.6	70.4	50	3	US-09-228-455-1	Sequence 2, App1
C 74	17.6	70.4	50	3	US-09-228-455-2	Sequence 17, App1
C 75	17.6	70.4	50	3	US-09-276-533A-17	Sequence 17, App1
C 76	17.6	70.4	50	3	US-10-158-735-17	Sequence 12, App1
C 77	17.6	70.4	50	3	US-09-443-282B-12	Sequence 12, App1
C 78	17.6	70.4	50	3	US-10-154-517-1	Sequence 1, App1
C 79	17.6	70.4	50	3	US-10-154-517-2	Sequence 1, App1
C 80	17.6	70.4	50	3	US-10-032-017-11	Sequence 11, App1
C 81	17.6	70.4	50	3	US-09-248-796A-9723	Sequence 9723, App
C 82	17.6	70.4	50	3	US-09-702-705-863	Sequence 863, App
C 83	17.6	70.4	50	3	US-09-614-125B-863	Sequence 863, App
C 84	17.6	70.4	50	3	US-09-671-824-863	Sequence 863, App
C 85	17.6	70.4	50	3	US-09-658-824-863	Sequence 863, App
C 86	17.6	70.4	50	3	US-10-017-754-863	Sequence 863, App
C 87	17.6	70.4	50	3	US-09-651-976-863	Sequence 863, App
C 88	17.6	70.4	50	3	US-09-621-976-17988	Sequence 17988, A
C 89	17.6	70.4	50	3	US-09-949-016-48834	Sequence 48834, A
C 90	17.6	70.4	50	3	US-09-949-016-49006	Sequence 49006, A
C 91	17.6	70.4	50	3	US-09-949-016-49178	Sequence 49178, A
C 92	17.6	70.4	50	3	US-09-949-016-49354	Sequence 49354, A
C 93	17.6	70.4	50	3		
C 94	17.6	70.4	50	3		
C 95	17.6	70.4	50	3		
C 96	17.6	70.4	50	3		
C 97	17.6	70.4	50	3		

```

C 98 17.6 70.4 601 3 US-09-949-016-52788 Sequence 52788, A
C 99 17.6 70.4 601 3 US-09-949-016-63223 Sequence 63223, A
C 100 17.6 70.4 601 3 US-09-949-016-87036 Sequence 87036, A
C 101 17.6 70.4 601 3 US-09-949-016-87037 Sequence 87037, A
C 102 17.6 70.4 601 3 US-09-949-016-87038 Sequence 87038, A
C 103 17.6 70.4 601 3 US-09-949-016-116552 Sequence 116552, A
C 104 17.6 70.4 601 3 US-09-949-016-135595 Sequence 135595, A
C 105 17.6 70.4 601 3 US-09-949-016-135596 Sequence 135596, A
C 106 17.6 70.4 601 3 US-09-949-016-182482 Sequence 182482, A
C 107 17.6 70.4 601 3 US-09-949-016-182483 Sequence 182483, A
C 108 17.6 70.4 601 3 US-09-949-016-182484 Sequence 182484, A
C 109 17.6 70.4 601 3 US-09-949-016-182602 Sequence 182602, A
C 110 17.6 70.4 601 3 US-09-949-016-182603 Sequence 182603, A
C 111 17.6 70.4 601 3 US-09-949-016-182604 Sequence 182604, A
C 112 17.6 70.4 601 3 US-09-949-016-184441 Sequence 184441, A
C 113 17.6 70.4 601 3 US-09-949-016-184442 Sequence 184442, A
C 114 17.6 70.4 669 3 US-09-328-111-342 Sequence 342, App
C 115 17.6 70.4 2309 3 US-09-091-725-9 Sequence 9, Appl
C 116 17.6 70.4 9139 3 US-09-322-478-22 Sequence 22, Appl
C 117 17.6 70.4 9139 3 US-09-586-106D-22 Sequence 22, Appl
C 118 17.6 70.4 9139 3 US-10-799-870-22 Sequence 22, Appl
C 119 17.6 70.4 9556 3 US-09-949-016-17303 Sequence 17303, A
C 120 17.6 70.4 15950 3 US-09-949-016-16383 Sequence 16383, A
C 121 17.6 70.4 25982 3 US-09-949-016-17308 Sequence 17308, A
C 122 17.6 70.4 26452 3 US-09-949-016-15822 Sequence 15822, A
C 123 17.6 70.4 27223 3 US-09-949-016-13036 Sequence 13036, A
C 124 17.6 70.4 32654 3 US-09-801-191A-3 Sequence 3, Appl
C 125 17.6 70.4 32654 3 US-10-345-198-3 Sequence 3, Appl
C 126 17.6 70.4 33529 3 US-09-949-016-12865 Sequence 12865, A
C 127 17.6 70.4 33529 3 US-09-949-016-17364 Sequence 17364, A
C 128 17.6 70.4 35725 3 US-09-949-016-13772 Sequence 13772, A
C 129 17.6 70.4 41863 3 US-09-949-016-14948 Sequence 14948, A
C 130 17.6 70.4 43307 3 US-09-949-016-15774 Sequence 15774, A
C 131 17.6 70.4 46323 3 US-09-949-016-12526 Sequence 12526, A
C 132 17.6 70.4 46323 3 US-09-949-016-13677 Sequence 13677, A
C 133 17.6 70.4 47030 3 US-09-949-016-13037 Sequence 13037, A
C 134 17.6 70.4 47030 3 US-09-949-016-13038 Sequence 13038, A
C 135 17.6 70.4 47030 3 US-09-949-016-15039 Sequence 15039, A
C 136 17.6 70.4 47030 3 US-09-949-016-15039 Sequence 15039, A
C 137 17.6 70.4 55280 3 US-09-949-016-15040 Sequence 15040, A
C 138 17.6 70.4 75378 3 US-09-949-016-17140 Sequence 17140, A
C 139 17.6 70.4 80355 3 US-09-949-016-12735 Sequence 12735, A
C 140 17.6 70.4 80357 3 US-09-949-016-13572 Sequence 13572, A
C 141 17.6 70.4 88450 3 US-09-949-016-12758 Sequence 12758, A
C 142 17.6 70.4 88736 3 US-09-949-016-14222 Sequence 14222, A
C 143 17.6 70.4 95561 3 US-09-949-016-12768 Sequence 12768, A
C 144 17.6 70.4 95561 3 US-09-949-016-13306 Sequence 13306, A
C 145 17.6 70.4 95561 3 US-09-949-016-13307 Sequence 13307, A
C 146 17.6 70.4 111454 3 US-09-949-016-14196 Sequence 14196, A
C 147 17.6 70.4 111459 3 US-09-949-016-12337 Sequence 12337, A
C 148 17.6 70.4 112874 3 US-09-949-016-13180 Sequence 13180, A
C 149 17.6 70.4 113042 3 US-09-949-016-12343 Sequence 12343, A
C 150 17.6 70.4 113042 3 US-09-949-016-15246 Sequence 15246, A

```

ALIGNMENTS

```

RESULT 1
US-09-949-016-33868
; Sequence 33868, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03

```

```

; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33868
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-33868

```

```

Query Match
Best Local Similarity 91.3%; Score 19.8; DB 3; Length 601;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 3 AAAAAAACTATGCTTATCTT 25
Db 64 AAAAAAACTATGCTTATCTT 86

```

```

RESULT 2
US-09-949-016-96073
; Sequence 96073, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231, 498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-96073

```

```

Query Match
Best Local Similarity 91.3%; Score 19.8; DB 3; Length 601;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 3 AAAAAAACTATGCTTATCTT 25
Db 64 AAAAAAACTATGCTTATCTT 86

```

```

RESULT 3
US-09-949-016-14543/C
; Sequence 14543, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231, 498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14543

```

```

; LENGTH: 387902
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(387902)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14543

Query Match
Best Local Similarity 91.3%; Score 19.8; DB 3; Length 387902;
Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAAAAAAACTATAGCTTGATCTT 25
Db 137296 AAAAAAAAACTATATCTTATCTT 137274

RESULT 4
US-09-949-016-12557/C
; Sequence 12557, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12557
; LENGTH: 421883
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(421883)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12557

Query Match
Best Local Similarity 91.3%; Score 19.8; DB 3; Length 421883;
Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAAAAAAACTATAGCTTGATCTT 25
Db 137296 AAAAAAAAACTATATCTTATCTT 137274

RESULT 5
US-09-270-767-5200
; Sequence 5200, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5200
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-5200

Query Match
Best Local Similarity 79.2%; Score 19.8; DB 3; Length 377;
Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 4 AAAAAAAAACTTATGTTGATTT 27

RESULT 6
US-09-270-767-20482
; Sequence 20482, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 20482
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-20482

Query Match
Best Local Similarity 87.5%; Score 19.2; DB 3; Length 377;
Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 4 AAAAAAAAACTTATGTTGATTT 27

RESULT 7
US-09-949-016-15095/C
; Sequence 15095, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15095
; LENGTH: 93510
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(93510)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15095

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 3; Length 93510;
Pred. No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 18619 AAAAAAAAACTTAACGTGATCT 18596
```

RESULT 8
US-08-858-207A-40
Sequence 40, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328el Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1117 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-40
Query Match 76.0%; Score 19; DB 3; Length 1117;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AAAAAAAAACTATAGCTTG 20
DB 1014 AAAAAAAAACTATAGCTTG 1032
RESULT 9
US-09-248-796A-638
Sequence 638, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13

NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 638
LENGTH: 2727
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-638
Query Match 75.2%; Score 18.8; DB 3; Length 2727;
Best Local Similarity 90.9%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 AAAAAAAAACTATAGCTTGATCT 24
DB 7 AAAAAAAAACTATAGCTTGATCT 28
RESULT 10
US-09-949-016-16469/C
Sequence 16469, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16469
LENGTH: 18264
TYPE: DNA
ORGANISM: Human
US-09-949-016-16469
Query Match 75.2%; Score 18.8; DB 3; Length 18264;
Best Local Similarity 90.9%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTTGAT 22
DB 17376 AAAAAAAAACTATAGCAGAT 17355
RESULT 11
US-09-949-016-15321/C
Sequence 15321, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15321
LENGTH: 74790
TYPE: DNA
ORGANISM: Human

US-09-949-016-15321

Query Match 75.2%; Score 18.8; DB 3; Length 74790;
Best Local Similarity 90.9%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGAT 22
Db 31762 AAAAAAAAACTAGCTTAAT 31741

RESULT 12

US-09-949-016-20819/c
; Sequence 20819, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20819
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-20819

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 173 AAAAAAAAAATTATGCTTGACCT 149

RESULT 13

US-09-949-016-67144/c
; Sequence 67144, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67144
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-67144

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 173 AAAAAAAAAATTATGCTTGACCT 149

RESULT 14

US-09-949-016-85277/c
; Sequence 85277, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85277
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-85277

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 173 AAAAAAAAAATTATGCTTGACCT 149

RESULT 15

US-09-949-016-86539/c
; Sequence 86539, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86539
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-86539

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 540 AAAAAAAAACTAGCAACATCTT 516

```
RESULT 16
US-09-270-767-9667/c
; Sequence 9667, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9667
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9667

Query Match          74.4%; Score 18.6; DB 3; Length 627;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 122 AAAAAAAAACTAAGCTCGATATT 98

RESULT 17
US-09-270-767-24949/c
; Sequence 24949, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24949
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24949

Query Match          74.4%; Score 18.6; DB 3; Length 627;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 122 AAAAAAAAACTAAGCTCGATATT 98

RESULT 18
US-09-949-016-14171
; Sequence 14171, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14171
; LENGTH: 8829
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14171

Query Match          74.4%; Score 18.6; DB 3; Length 8829;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 6384 AAAAAAAAAATTATGCTTGACCTT 6408

RESULT 19
US-09-693-205A-9/c
; Sequence 9, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; METHODS OF USE THEREFOR
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; CURRENT FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 12717
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-693-205A-9

Query Match          74.4%; Score 18.6; DB 3; Length 12717;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 4520 AAGACCAACTATATCTTGATGTT 4496

RESULT 20
US-09-693-205A-13/c
; Sequence 13, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; METHODS OF USE THEREFOR
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; CURRENT FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 12717
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-693-205A-13

Query Match          74.4%; Score 18.6; DB 3; Length 12717;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
```

```
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 4520 AAAAGACAACTATATCTTGATGTT 4496

RESULT 21
US-09-693-205A-14/c
; Sequence 14, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 12717
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-693-205A-14

Query Match 74.4%; Score 18.6; DB 3; Length 12717;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 4520 AAAAGACAACTATATCTTGATGTT 4496

RESULT 22
US-09-693-205A-7/c
; Sequence 7, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 12792
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-693-205A-7

Query Match 74.4%; Score 18.6; DB 3; Length 12792;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 4596 AAAAGACAACTATATCTTGATGTT 4572

RESULT 23
US-09-693-205A-1/c
```

```
; Sequence 1, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 12793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-693-205A-1

Query Match 74.4%; Score 18.6; DB 3; Length 12793;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 4596 AAAAGACAACTATATCTTGATGTT 4572

RESULT 24
US-09-693-205A-11/c
; Sequence 11, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 12793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-693-205A-11

Query Match 74.4%; Score 18.6; DB 3; Length 12793;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 4596 AAAAGACAACTATATCTTGATGTT 4572

RESULT 25
US-09-693-205A-12/c
; Sequence 12, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; FILE REFERENCE: 2825.1021-003
```

```
/ CURRENT APPLICATION NUMBER: US/09/693,205A
/ CURRENT FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: US 60/160,588
/ PRIOR FILING DATE: 1999-10-20
/ NUMBER OF SEQ ID NOS: 73
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 12
/ LENGTH: 12793
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-693-205A-12

Query Match          74.4%; Score 18.6; DB 3; Length 12793;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 4596 AAAAAAGCAAACTATATCTTGATGTT 4572

RESULT 26
US-09-693-205A-15/c
/ Sequence 15, Application US/09693205A
/ Patent No. 6812333
/ GENERAL INFORMATION:
/ APPLICANT: Hudson, Thomas J.
/ APPLICANT: Engert, James C.
/ APPLICANT: Richter, Andrea
/ TITLE OF INVENTION: IDENTIFICATION OF ARSAS MUTATIONS AND
/ FILE REFERENCE: 2825.1021-003
/ CURRENT APPLICATION NUMBER: US/09/693,205A
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: US 60/160,588
/ PRIOR FILING DATE: 1999-10-20
/ NUMBER OF SEQ ID NOS: 73
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15
/ LENGTH: 12793
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-693-205A-15

Query Match          74.4%; Score 18.6; DB 3; Length 12793;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 4596 AAAAAAGCAAACTATATCTTGATGTT 4572

RESULT 27
US-09-949-016-11895
/ Sequence 10, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11895
```

```
/ LENGTH: 13115
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-11895

Query Match          74.4%; Score 18.6; DB 3; Length 13115;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 6382 AAAAAAAAAATATGCTTGACCT 6406

RESULT 28
US-09-949-016-13723
/ Sequence 13723, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13723
/ LENGTH: 13371
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-13723

Query Match          74.4%; Score 18.6; DB 3; Length 13371;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 6384 AAAAAAAAAATATGCTTGACCT 6408

RESULT 29
US-09-817-762-10
/ Sequence 10, Application US/09817762
/ Patent No. 6858774
/ GENERAL INFORMATION:
/ APPLICANT: Spalding, Edgar P.
/ APPLICANT: No. 6858774, Bostl
/ TITLE OF INVENTION: MDR-Like ABC Transporter Gene From
/ FILE REFERENCE: 13238-00061
/ CURRENT APPLICATION NUMBER: US/09/817,762
/ PRIOR FILING DATE: 2001-03-26
/ PRIOR APPLICATION NUMBER: PCT/US99/22363
/ PRIOR FILING DATE: 1999-09-24
/ PRIOR APPLICATION NUMBER: US 60/101,814
/ PRIOR FILING DATE: 1998-09-25
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10
/ LENGTH: 14000
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (3425)...(0)
```

```
OTHER INFORMATION: Translation start codon
;
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Genbank AP000386
; DATABASE ENTRY DATE: 1999-08-03
US-09-817-762-10
```

```
Query Match          74.4% Score 18.6; DB 3; Length 14000;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 5012 AGAAAAAGACTATAGATTCATTTT 5036
```

```
RESULT 30
US-09-949-016-15821/c
; Sequence 15821, Application US/09949016
; Patent No. 6812339
```

```
GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```

```
PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 15821
```

```
LENGTH: 50563
```

```
TYPE: DNA
```

```
ORGANISM: Human
```

```
US-09-949-016-15821
```

```
Query Match
```

```
Best Local Similarity 74.4% Score 18.6; DB 3; Length 50563;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 35461 AAAAAAGTATTAATTCATTTT 35437
```

```
RESULT 31
US-09-949-016-15433
; Sequence 15433, Application US/09949016
; Patent No. 6812339
```

```
GENERAL INFORMATION:
```

```
APPLICANT: VENTER, J. Craig et al.
```

```
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
FILE REFERENCE: CL001307
```

```
CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
```

```
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```

```
PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 15433
```

```
LENGTH: 51849
```

```
TYPE: DNA
```

```
ORGANISM: Human
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
; LOCATION: (1)...(51849)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15433
```

```
Query Match          74.4% Score 18.6; DB 3; Length 51849;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 40005 AAAAAATTAATTAATTCATTTT 40029
```

```
RESULT 32
US-09-949-016-12461
; Sequence 12461, Application US/09949016
; Patent No. 6812339
```

```
GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```

```
PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 12461
```

```
LENGTH: 93920
```

```
TYPE: DNA
```

```
ORGANISM: Human
```

```
US-09-949-016-12461
```

```
Query Match
```

```
Best Local Similarity 74.4% Score 18.6; DB 3; Length 93920;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 78341 AAAAAAAACCATTCATTCATCTT 78365
```

```
RESULT 33
US-09-949-016-16853
; Sequence 16853, Application US/09949016
; Patent No. 6812339
```

```
GENERAL INFORMATION:
```

```
APPLICANT: VENTER, J. Craig et al.
```

```
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
FILE REFERENCE: CL001307
```

```
CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
```

```
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```

```
PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 16853
```

```
LENGTH: 93920
```

```
TYPE: DNA
```

```
ORGANISM: Human
```

```
US-09-949-016-16853
```


US-09-949-016-13539

Query Match 74.4%; Score 18.6; DB 3; Length 314798;
Best Local Similarity 84.0%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTGATCTT 25
DB 95668 AAAAAAAAACTATAGCTTACTT 95644

RESULT 38

US-09-949-016-11769
; Sequence 11769, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11769
; LENGTH: 105919
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(105919)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11769

Query Match 73.6%; Score 18.4; DB 3; Length 105919;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTG 20
DB 15236 AAAAAAAAACTAGCTTG 15255

RESULT 39

US-09-248-796A-10034
; Sequence 10034, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 10034
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-10034

Query Match 72.8%; Score 18.2; DB 3; Length 246;
Best Local Similarity 87.0%; Pred. No. 5.7e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCT 24
DB 90 AAAAAAAAACTATAGCTTGAGCT 112

RESULT 40

US-09-949-016-19777/c
; Sequence 19777, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19777
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-19777

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 5.7e+02;
Matches 20; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 310 AAAAAAAAAATATGCTAGAACTT 286

RESULT 41

US-09-949-016-20890
; Sequence 20890, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20890
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-20890

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 80.0%; Pred. No. 5.7e+02;
Matches 20; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 300 AAAAAAAAACTATAGCTTGCTATT 324

```
RESULT 42
US-09-543-681A-4161/c
; Sequence 4161, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4161
; LENGTH: 945
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-4161

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 3; Length 945;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATC 23
DB 653 AAAAAAAAACTATAGCTTGATC 631

RESULT 43
US-09-270-767-9472/c
; Sequence 9472, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9472
; LENGTH: 1184
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9472

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 3; Length 1184;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATC 23
DB 61 AAAAAAAAACTATAGCTTGATC 39

RESULT 44
US-09-270-767-24754/c
; Sequence 24754, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24754
; LENGTH: 1184
; TYPE: DNA
```

```
; ORGANISM: Drosophila melanogaster
US-09-270-767-24754

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 3; Length 1184;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATC 23
DB 61 AAAAAAAAACTATAGCTTGATC 39

RESULT 45
US-09-949-002-693/c
; Sequence 693, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 693
; LENGTH: 63467
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-693

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 3; Length 63467;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATC 23
DB 35713 AAAAAAAAACTATAGATTTTTC 35691

RESULT 46
US-09-949-016-17289/c
; Sequence 17289, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17289
; LENGTH: 64518
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17289

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 3; Length 64518;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATC 23
```



```
Db      35713 AAAAAAAAACTATAGATTTC 35691
|||||
RESULT 47
US-09-949-002-843/c
; Sequence 843, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: C1000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 843
; LENGTH: 64518
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-843

Query Match      72.8%; Score 18.2; DB 3; Length 64518;
Best Local Similarity 87.0%; Pred. No. 5.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTATAGCTTGATC 23
|||||
Db      35713 AAAAAAAAACTATAGATTTC 35691
|||||

RESULT 48
US-09-949-016-15630/c
; Sequence 15630, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15630
; LENGTH: 112705
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112705)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15630

Query Match      72.8%; Score 18.2; DB 3; Length 112705;
Best Local Similarity 87.0%; Pred. No. 5.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTATAGCTTGATC 23
|||||
Db      11586 AAAAAATATCTATAGCTGAGC 11564
|||||
```

```
RESULT 49
US-09-248-796A-13477
; Sequence 13477, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinscock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 13477
; LENGTH: 234
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-13477

Query Match      71.2%; Score 17.8; DB 3; Length 234;
Best Local Similarity 90.5%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 AAAAAAAAACTATAGCTTGATCT 24
|||||
Db      50 AAAAAAAAAATAGCTTGATCT 70
|||||

RESULT 50
US-09-621-976-17000/c
; Sequence 17000, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: ESTE and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 17000
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-17000

Query Match      71.2%; Score 17.8; DB 3; Length 372;
Best Local Similarity 90.5%; Pred. No. 8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTATAGCTTGA 21
|||||
Db      364 AAAAAAAAAACCATTCCTTGA 344
|||||

Search completed: December 14, 2005, 07:44:26
Job time : 61.5 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds
(without alignments)
800,703 Million cell updates/sec

Title: US-10-681-773-5
Perfect score: 25
Sequence: 1 aaaaaaaaaagcatgactgacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/1.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/H.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PCBUS.COMB.seq: *
7: /cgn2_6/ptodata/1/ina/RE.COMB.seq: *
8: /cgn2_6/ptodata/1/ina/RE.COMB.seq: *
9: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
No.						
C 1	19.2	76.8	1139	3	US-08-755-587-40	Sequence 40, Appl
C 2	19.2	76.8	7699	3	US-09-949-016-2989	Sequence 2989, Ap
C 3	19.2	76.8	17930	3	US-09-949-016-11902	Sequence 11902, A
C 4	19.2	76.8	18351	3	US-09-949-016-16749	Sequence 16749, A
C 5	19.2	76.8	26843	3	US-09-949-016-17208	Sequence 17208, A
C 6	19.2	76.8	103894	3	US-09-949-016-14450	Sequence 14450, A
C 7	19.2	76.8	104520	3	US-09-949-016-13303	Sequence 13303, A
C 8	19.2	76.8	126029	3	US-09-949-016-14731	Sequence 14731, A
C 9	19.2	76.8	168104	3	US-09-949-016-12026	Sequence 12026, A
C 10	19.2	76.8	168105	3	US-09-949-016-16554	Sequence 16554, A
C 11	18.8	75.2	601	3	US-09-949-016-170298	Sequence 170298, A
C 12	18.8	75.2	18474	3	US-09-949-016-17559	Sequence 17559, A
C 13	18.8	75.2	87617	3	US-09-949-016-16551	Sequence 16551, A
C 14	18.6	74.4	601	3	US-09-949-016-15339	Sequence 15339, A
C 15	18.6	74.4	20560	3	US-09-949-016-88723	Sequence 88723, A
C 16	18.6	74.4	40493	3	US-09-949-016-12719	Sequence 12719, A
C 17	18.6	74.4	40493	3	US-09-949-016-15453	Sequence 15453, A
C 18	18.6	74.4	58361	3	US-09-949-016-16755	Sequence 16755, A
C 19	18.6	74.4	58361	3	US-09-949-016-16756	Sequence 16756, A
C 20	18.6	74.4	59065	3	US-09-813-817-3	Sequence 3, Appl1
C 21	18.6	74.4	59065	3	US-09-978-197-3	Sequence 3, Appl1
C 22	18.6	74.4	59065	3	US-10-135-656-3	Sequence 3, Appl1
C 23	18.6	74.4	59065	3	US-10-820-230-3	Sequence 3, Appl1
C 24	18.6	74.4	71815	3	US-09-949-016-12501	Sequence 12501, A

C 25	18.6	74.4	93971	3	US-09-949-016-16097	Sequence 16097, A
C 26	18.6	74.4	93971	3	US-09-949-016-16098	Sequence 16098, A
C 27	18.6	74.4	147321	3	US-09-949-016-15450	Sequence 15450, A
C 28	18.6	74.4	150597	3	US-09-949-016-15379	Sequence 15379, A
C 29	18.6	74.4	160552	3	US-09-593-828-11	Sequence 11, Appl
C 30	18.6	74.4	174259	3	US-09-949-016-11968	Sequence 11968, A
C 31	18.6	74.4	174262	3	US-09-949-016-14259	Sequence 14259, A
C 32	18.6	74.4	421118	3	US-09-949-016-16297	Sequence 16297, A
C 33	18.4	73.6	264206	3	US-09-949-016-12711	Sequence 12711, A
C 34	18.4	73.6	264304	3	US-09-949-016-13249	Sequence 13249, A
C 35	18.2	72.8	601	3	US-10-013-598-3	Sequence 3, Appl1
C 36	18.2	72.8	601	3	US-09-949-016-17634	Sequence 17634, A
C 37	18.2	72.8	601	3	US-09-949-016-26597	Sequence 26597, A
C 38	18.2	72.8	601	3	US-09-949-016-16198	Sequence 36198, A
C 39	18.2	72.8	601	3	US-09-949-016-54606	Sequence 54606, A
C 40	18.2	72.8	601	3	US-09-949-016-54638	Sequence 54638, A
C 41	18.2	72.8	601	3	US-09-949-016-54670	Sequence 54670, A
C 42	18.2	72.8	601	3	US-09-949-016-66376	Sequence 66376, A
C 43	18.2	72.8	601	3	US-09-949-016-119035	Sequence 119035, A
C 44	18.2	72.8	601	3	US-09-949-016-119071	Sequence 119071, A
C 45	18.2	72.8	601	3	US-09-949-016-119107	Sequence 119107, A
C 46	18.2	72.8	601	3	US-09-949-016-119143	Sequence 119143, A
C 47	18.2	72.8	601	3	US-09-949-016-119179	Sequence 119179, A
C 48	18.2	72.8	601	3	US-09-949-016-119215	Sequence 119215, A
C 49	18.2	72.8	601	3	US-09-949-016-142686	Sequence 142686, A
C 50	18.2	72.8	601	3	US-09-949-016-152429	Sequence 152429, A
C 51	18.2	72.8	601	3	US-09-949-016-196001	Sequence 196001, A
C 52	18.2	72.8	12118	3	US-09-788-654A-3	Sequence 3, Appl1
C 53	18.2	72.8	16885	3	US-09-949-016-17297	Sequence 17297, A
C 54	18.2	72.8	16890	3	US-09-949-016-12350	Sequence 12190, A
C 55	18.2	72.8	19451	3	US-09-949-016-13695	Sequence 13695, A
C 56	18.2	72.8	48185	3	US-09-949-016-17027	Sequence 17027, A
C 57	18.2	72.8	74804	3	US-09-949-016-15118	Sequence 15118, A
C 58	18.2	72.8	75212	3	US-09-949-016-13313	Sequence 13313, A
C 59	18.2	72.8	75212	3	US-09-949-016-13315	Sequence 13315, A
C 60	18.2	72.8	75212	3	US-09-949-016-13315	Sequence 13315, A
C 61	18.2	72.8	77994	3	US-09-949-016-16201	Sequence 12517, A
C 62	18.2	72.8	77994	3	US-09-949-016-16201	Sequence 16201, A
C 63	18.2	72.8	80004	3	US-09-949-016-16318	Sequence 16318, A
C 64	18.2	72.8	87523	3	US-09-949-016-12670	Sequence 12670, A
C 65	18.2	72.8	87523	3	US-09-949-016-15047	Sequence 15047, A
C 66	18.2	72.8	87523	3	US-09-949-016-15048	Sequence 15048, A
C 67	18.2	72.8	87523	3	US-09-949-016-15048	Sequence 15048, A
C 68	18.2	72.8	87869	3	US-09-949-016-15049	Sequence 15049, A
C 69	18.2	72.8	87869	3	US-09-949-016-15044	Sequence 15044, A
C 70	18.2	72.8	87869	3	US-09-949-016-15045	Sequence 15045, A
C 71	18.2	72.8	87869	3	US-09-949-016-15046	Sequence 15046, A
C 72	18.2	72.8	101894	3	US-09-949-016-12005	Sequence 12005, A
C 73	18.2	72.8	117937	3	US-09-949-016-12762	Sequence 12762, A
C 74	18.2	72.8	117937	3	US-09-949-016-12762	Sequence 12762, A
C 75	18.2	72.8	134292	3	US-09-949-016-12158	Sequence 12158, A
C 76	18.2	72.8	146039	3	US-09-949-016-12449	Sequence 12449, A
C 77	18.2	72.8	152123	3	US-09-949-016-13845	Sequence 13845, A
C 78	18.2	72.8	152145	3	US-09-949-016-12371	Sequence 12371, A
C 79	18.2	72.8	346112	3	US-09-949-016-13165	Sequence 13165, A
C 80	18.2	72.8	1664976	3	US-08-916-421B-1	Sequence 1, Appl1
C 81	18.2	72.8	1830121	3	US-09-692-570-1	Sequence 1, Appl1
C 82	18.2	72.8	1830121	3	US-09-692-570-1	Sequence 1, Appl1
C 83	18.2	72.8	1830121	3	US-09-643-990A-1	Sequence 1, Appl1
C 84	18.2	72.8	1830121	3	US-10-158-865-1	Sequence 1, Appl1
C 85	17.8	71.2	601	3	US-09-949-016-61678	Sequence 1, Appl1
C 86	17.8	71.2	601	3	US-09-949-016-61678	Sequence 16178, A
C 87	17.8	71.2	601	3	US-09-949-016-61800	Sequence 61800, A
C 88	17.8	71.2	601	3	US-09-949-016-61800	Sequence 61801, A
C 89	17.8	71.2	601	3	US-09-949-016-61800	Sequence 61801, A
C 90	17.8	71.2	11048	3	US-09-949-016-173429	Sequence 173429, A
C 91	17.8	71.2	55806	3	US-09-949-016-15605	Sequence 17018, A
C 92	17.8	71.2	85854	3	US-09-949-016-15605	Sequence 15605, A
C 93	17.8	71.2	145287	3	US-09-949-016-135308	Sequence 12908, A
C 94	17.8	71.2	145287	3	US-09-949-016-135308	Sequence 13530, A
C 95	17.8	71.2	219964	3	US-09-949-016-15086	Sequence 13531, A
C 96	17.6	70.4	472	3	US-09-142-078-55	Sequence 15066, A
C 97	17.6	70.4	472	3	US-09-357-141-55	Sequence 55, Appl1

C 98 17.6 70.4 472 3 US-09-533-889-55 Sequence 55, Appl
C 99 17.6 70.4 472 3 US-09-142-080-55 Sequence 55, Appl
C 100 17.6 70.4 601 3 US-09-949-016-18060 Sequence 18060, A
C 101 17.6 70.4 601 3 US-09-949-016-26215 Sequence 26215, A
C 102 17.6 70.4 601 3 US-09-949-016-44741 Sequence 44741, A
C 103 17.6 70.4 601 3 US-09-949-016-44742 Sequence 44742, A
C 104 17.6 70.4 601 3 US-09-949-016-51161 Sequence 51161, A
C 105 17.6 70.4 601 3 US-09-949-016-59069 Sequence 59069, A
C 106 17.6 70.4 601 3 US-09-949-016-59070 Sequence 59070, A
C 107 17.6 70.4 601 3 US-09-949-016-74117 Sequence 74117, A
C 108 17.6 70.4 601 3 US-09-949-016-74148 Sequence 74148, A
C 109 17.6 70.4 601 3 US-09-949-016-80327 Sequence 80327, A
C 110 17.6 70.4 601 3 US-09-949-016-80328 Sequence 80328, A
C 111 17.6 70.4 601 3 US-09-949-016-80329 Sequence 80329, A
C 112 17.6 70.4 601 3 US-09-949-016-88783 Sequence 88783, A
C 113 17.6 70.4 601 3 US-09-949-016-134264 Sequence 134264, A
C 114 17.6 70.4 601 3 US-09-949-016-140887 Sequence 140887, A
C 115 17.6 70.4 601 3 US-09-949-016-143240 Sequence 143240, A
C 116 17.6 70.4 601 3 US-09-949-016-143241 Sequence 143241, A
C 117 17.6 70.4 601 3 US-09-949-016-143411 Sequence 143411, A
C 118 17.6 70.4 601 3 US-09-949-016-143412 Sequence 143412, A
C 119 17.6 70.4 601 3 US-09-949-016-158316 Sequence 158316, A
C 120 17.6 70.4 601 3 US-09-949-016-171633 Sequence 171633, A
C 121 17.6 70.4 601 3 US-09-949-016-183396 Sequence 183396, A
C 122 17.6 70.4 601 3 US-09-949-016-183422 Sequence 183422, A
C 123 17.6 70.4 601 3 US-09-949-016-189423 Sequence 189423, A
C 124 17.6 70.4 601 3 US-09-949-016-189424 Sequence 189424, A
C 125 17.6 70.4 601 3 US-09-949-016-189425 Sequence 189425, A
C 126 17.6 70.4 601 3 US-09-949-016-190271 Sequence 190271, A
C 127 17.6 70.4 601 3 US-09-949-016-190272 Sequence 190272, A
C 128 17.6 70.4 601 3 US-09-949-016-197087 Sequence 197087, A
C 129 17.6 70.4 601 3 US-09-949-016-198074 Sequence 198074, A
C 130 17.6 70.4 782 3 US-09-328-475C-144 Sequence 1404, A
C 131 17.6 70.4 4370 3 US-09-976-594-885 Sequence 885, App
C 132 17.6 70.4 4529 3 US-09-949-016-14004 Sequence 14004, A
C 133 17.6 70.4 7505 3 US-09-078-294-13 Sequence 13, Appl
C 134 17.6 70.4 8050 3 US-09-491-362-11 Sequence 11, Appl
C 135 17.6 70.4 8050 3 US-09-874-562-11 Sequence 11, Appl
C 136 17.6 70.4 9226 3 US-09-949-016-12596 Sequence 12596, A
C 137 17.6 70.4 9226 3 US-09-949-016-16007 Sequence 16007, A
C 138 17.6 70.4 9997 3 US-09-949-016-12919 Sequence 12919, A
C 139 17.6 70.4 11406 3 US-09-949-016-17555 Sequence 17555, A
C 140 17.6 70.4 15766 3 US-09-338-807-73 Sequence 73, Appl
C 141 17.6 70.4 15766 3 US-09-218-207-73 Sequence 73, Appl
C 142 17.6 70.4 16573 3 US-09-949-016-11764 Sequence 11764, A
C 143 17.6 70.4 17370 3 US-09-949-016-17331 Sequence 17331, A
C 144 17.6 70.4 18744 3 US-09-949-016-14052 Sequence 14052, A
C 145 17.6 70.4 24020 3 US-09-949-016-17353 Sequence 17353, A
C 146 17.6 70.4 25656 3 US-09-949-016-13022 Sequence 13022, A
C 147 17.6 70.4 27663 3 US-09-949-016-16160 Sequence 16160, A
C 148 17.6 70.4 31769 3 US-09-949-002-734 Sequence 734, App
C 149 17.6 70.4 35840 3 US-09-949-016-12156 Sequence 12156, A
C 150 17.6 70.4 35840 3 US-09-949-016-13907 Sequence 13907, A

ALIGNMENTS

RESULT 1
US-08-755-587-40/c
Sequence 40, Application US/08755587
Patent No. 604597
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson

STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1139 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 501..639
FEATURE:
NAME/KEY: CDS
LOCATION: 501..639
US-08-755-587-40

Query Match 76.8%; Score 19.2; DB 3; Length 1139;
Best local similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 1038 AAAAAAAAAAGCTTGACTGTGAGA 1015

RESULT 2
US-09-949-016-2989/c
Sequence 2989, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCES: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2989
LENGTH: 7699
TYPE: DNA
ORGANISM: Human
US-09-949-016-2989

```
Query Match      76.8%; Score 19.2; DB 3; Length 7699;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAAAAAAAGCATGCTGTGACAC 25
Db      6510 AAAAAAAGCAGCAGCTTTGACAC 6487

RESULT 3
US-09-949-016-11902/c
; Sequence 11902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11902
; LENGTH: 17930
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11902

Query Match      76.8%; Score 19.2; DB 3; Length 17930;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGCTGTGACA 24
Db      14964 AAAAAAAGCATTACTATGACA 14941

RESULT 4
US-09-949-016-16749/c
; Sequence 16749, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16749
; LENGTH: 18351
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16749

Query Match      76.8%; Score 19.2; DB 3; Length 18351;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGCTGTGACA 24
Db      14964 AAAAAAAGCATTACTATGACA 14941

RESULT 5
US-09-949-016-17208
; Sequence 17208, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 17208
; LENGTH: 26843
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17208

Query Match      76.8%; Score 19.2; DB 3; Length 26843;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGCTGTGACA 24
Db      13603 AAAAAAAGACTGACTGTGACA 13626

RESULT 6
US-09-949-016-14450
; Sequence 14450, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 14450
; LENGTH: 103894
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(103894)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14450

Query Match      76.8%; Score 19.2; DB 3; Length 103894;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Db 30529 AAAAAAAAAAGTATGATGTGACA 30552

RESULT 7
US-09-949-016-13303/C
Sequence 13303, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13303
LENGTH: 104520
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(104520)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13303

Query Match 76.8%; Score 19.2; DB 3; Length 104520;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGTGTGAC 25
DB 98721 AAAAAAAGCAGCAGCTTGACAC 98698

RESULT 8
US-09-949-016-14731/C
Sequence 14731, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14731
LENGTH: 126029
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(126029)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14731

Query Match 76.8%; Score 19.2; DB 3; Length 126029;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGCATGTGTGAC 25
DB 122839 AAAAAAAGCAGCAGCTTGACAC 122816

RESULT 9
US-09-949-016-12026
Sequence 12026, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12026
LENGTH: 168104
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(168104)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12026

Query Match 76.8%; Score 19.2; DB 3; Length 168104;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTGTGACA 24
DB 74866 AAAAAAAGCATGTGTGCA 74889

RESULT 10
US-09-949-016-16554
Sequence 16554, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16554
LENGTH: 168105
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(168105)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16554

Query Match 76.8%; Score 19.2; DB 3; Length 168105;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGCA 24
Db 74866 AAAAAAAAAAGCATGACTGTGCA 74889

RESULT 11
US-09-949-016-170298/c
; Sequence 170298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 170298
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170298

Query Match 75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGCA 22
Db 171 AAAAAAAAAAGCATGACTGTGCA 150

RESULT 12
US-09-949-016-17559/c
; Sequence 17559, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 17559
; LENGTH: 18474
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17559

Query Match 75.2%; Score 18.8; DB 3; Length 18474;
Best Local Similarity 90.9%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGCA 22
Db 10955 AAAAAAAAAAGCATGACTGTGCA 10934

RESULT 13
US-09-949-016-16551/c
; Sequence 16551, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16551
; LENGTH: 87617
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(87617)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16551

Query Match 75.2%; Score 18.8; DB 3; Length 87617;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGCA 22
Db 21582 AAAAAAAAAAGCATGACTGTGCA 21561

RESULT 14
US-09-248-796A-11539
; Sequence 11539, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 11539
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-11539

Query Match 74.4%; Score 18.6; DB 3; Length 261;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 133 AAAAAAAAAAGCATGACTGTGACAC 157

```
RESULT 15
US-09-949-016-88723
; Sequence 88723, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 88723
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-88723

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 292 AAAAAAAAAAGTTTACTGTGGACAC 316

RESULT 16
US-09-949-016-12719
; Sequence 12719, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12719
; LENGTH: 20560
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(20560)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12719

Query Match          74.4%; Score 18.6; DB 3; Length 20560;
Best Local Similarity 84.0%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
RESULT 17
US-09-949-016-15453
; Sequence 15453, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15453
; LENGTH: 40493
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(40493)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15453

Query Match          74.4%; Score 18.6; DB 3; Length 40493;
Best Local Similarity 84.0%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 13660 AAAAAAAAAATCATGTAAGTGATAC 13684

RESULT 18
US-09-949-016-16755
; Sequence 16755, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16755
; LENGTH: 58361
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16755
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 58361;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 10624 ACIAAAAAAAAACTGTGCTGTGAC 10648
```

RESULT 19


```
US-09-949-016-16756
; Sequence 16756, Application US/09949016
; Patent No. 6812319
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16756
; LENGTH: 58361
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16756

Query Match          74.4%; Score 18.6; DB 3; Length 58361;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTGACAC 25
Db 10624 ACAAACAACTTGACTGTGTCTC 10648

RESULT 20
US-09-813-817-3
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1001178
; CURRENT APPLICATION NUMBER: US/09/813,817
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-813-817-3

Query Match          74.4%; Score 18.6; DB 3; Length 59065;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTGACAC 25
Db 11968 ACAAACAACTTGACTGTGTCTC 11992

RESULT 21
US-09-978-197-3
; Sequence 3, Application US/09978197
; Patent No. 6403353
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1001178DIV
; CURRENT APPLICATION NUMBER: US/09/978,197

US-10-135-696-3
; Sequence 3, Application US/10135696
; Patent No. 6740513
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1001178DIV
; CURRENT APPLICATION NUMBER: US/10/135,696
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 09/813,817
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/978,197
; PRIOR FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-135-696-3

Query Match          74.4%; Score 18.6; DB 3; Length 59065;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTGACAC 25
Db 11968 ACAAACAACTTGACTGTGTCTC 11992

RESULT 23
US-10-820-230-3
; Sequence 3, Application US/10820230
; Patent No. 6946545
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1001178DIV-III
; CURRENT APPLICATION NUMBER: US/10/820,230
; CURRENT FILING DATE: 2004-04-08
; PRIOR APPLICATION NUMBER: 10/135,696
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 09/978,197
; PRIOR FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: 09/813,817
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
/ SEQ ID NO 3
/ LENGTH: 59065
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-820-230-3
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 59065;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGCATGCTGTGACAC 25
Db      11968 AAAAAAAAAAAGCTGTGTGTGAC 11992
```

```
RESULT 24
US-09-949-016-12501
/ Sequence 12501, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 12501
/ LENGTH: 71815
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-12501
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 71815;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGCATGCTGTGACAC 25
Db      24078 AAAAAAAAAAAGCTGTGTGTGAC 24102
```

```
RESULT 25
US-09-949-016-16097
/ Sequence 16097, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 16097
/ LENGTH: 93971
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
```

```
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(93971)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16097
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 93971;
Best Local Similarity 84.0%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGCATGCTGTGACAC 25
Db      82251 AAAAAAAAAAGCCTGTGTGTGACTC 82275
```

```
RESULT 26
US-09-949-016-16098
/ Sequence 16098, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 16098
/ LENGTH: 93971
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(93971)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16098
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 93971;
Best Local Similarity 84.0%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGCATGCTGTGACAC 25
Db      82251 AAAAAAAAAAGCCTGTGTGTGACTC 82275
```

```
RESULT 27
US-09-949-016-15450
/ Sequence 15450, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15450
/ LENGTH: 147321
```

```

; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(147321)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15450

Query Match      74.4%; Score 18.6; DB 3; Length 147321;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 54893 AAAAAAAAAAGCATGCTGTGACAC 54917

RESULT 28
US-09-949-016-15379/C
; Sequence 15379, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15379
; LENGTH: 150597
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(150597)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15379

Query Match      74.4%; Score 18.6; DB 3; Length 150597;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 59104 AAAAAAAAAAGCATGCTGTGACAC 59080

RESULT 29
US-09-593-828-11/C
; Sequence 11, Application US/09593828
; Patent No. 6852518
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: No. 6852518el Glycosyl Sulfotransferases
; FILE REFERENCE: 6510-138US1
; CURRENT APPLICATION NUMBER: US/09/593,828
; CURRENT FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 160552
; TYPE: DNA

; ORGANISM: homo sapiens
US-09-593-828-11

Query Match      74.4%; Score 18.6; DB 3; Length 160552;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 140431 AAAAAAAAAAGCCTTATTGAGACAC 140407

RESULT 30
US-09-949-016-11968/C
; Sequence 11968, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11968
; LENGTH: 174259
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11968

Query Match      74.4%; Score 18.6; DB 3; Length 174259;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 142144 AAAAAAAAAAGTTTACTGGACAC 142120

RESULT 31
US-09-949-016-14259/C
; Sequence 14259, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14259
; LENGTH: 174262
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14259

Query Match      74.4%; Score 18.6; DB 3; Length 174262;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
```

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGAC 25
Db 142144 AAAAAAAAAAGTTTACTGGACAC 142120

RESULT 32
US-09-949-016-16297/C
; Sequence 16297, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16297
; LENGTH: 421118
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(421118)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16297

Query Match 74.4%; Score 18.6; DB 3; Length 421118;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGAC 25
Db 28193 AAAAAAAAAAGCATGACAC 28169

RESULT 33
US-09-949-016-12731/C
; Sequence 12731, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12731
; LENGTH: 264206
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12731

Query Match 73.6%; Score 18.4; DB 3; Length 264206;
Best Local Similarity 95.0%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGT 20
Db 23123 AAAAAAAAAAGCATGCTTT 23104

RESULT 34
US-09-949-016-13249/C
; Sequence 13249, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13249
; LENGTH: 264304
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13249

Query Match 73.6%; Score 18.4; DB 3; Length 264304;
Best Local Similarity 95.0%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGT 20
Db 23123 AAAAAAAAAAGCATGCTTT 23104

RESULT 35
US-10-013-598-3/C
; Sequence 3, Application US/10013598
; Patent No. 6872529
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc.
; APPLICANT: Su, Xing
; TITLE OF INVENTION: Complexity Management of Genomic DNA
; FILE REFERENCE: 3422
; CURRENT APPLICATION NUMBER: US/10/013,598
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 41
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic oligonucleotide
US-10-013-598-3

Query Match 72.8%; Score 18.2; DB 3; Length 41;
Best Local Similarity 87.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGAC 23
Db 37 AAAAAAAAAAATGACAGTAC 15

RESULT 36
US-09-949-016-17634
; Sequence 17634, Application US/09949016

```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17634
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17634

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

RESULT 37
US-09-949-016-26597
; Sequence 26597, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26597
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-26597

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGACTGTGAC 24
Db 534 AAAAAAAGAGCATGACTGTGACA 556

RESULT 38
US-09-949-016-36198
; Sequence 36198, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36198
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-36198

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

RESULT 39
US-09-949-016-54606
; Sequence 54606, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54606
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54606

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

RESULT 40
US-09-949-016-54638
; Sequence 54638, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54638
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54638
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 289 AAAAAAAAAAATTACTGTGAC 311
```

```
RESULT 41
US-09-949-016-54670
; Sequence 54670, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54670
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-54670
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 289 AAAAAAAAAAATTACTGTGAC 311
```

```
RESULT 42
US-09-949-016-66376
; Sequence 66376, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66376
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-66376
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 390 AAAAAAAAAAGCAGGCTGTGAC 412
```

```
RESULT 43
US-09-949-016-119035
; Sequence 119035, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119035
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119035
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 289 AAAAAAAAAAATTACTGTGAC 311
```

```
RESULT 44
US-09-949-016-119071
; Sequence 119071, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119071
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
```

US-09-949-016-1191071

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

RESULT 45

US-09-949-016-119107
Sequence 119107, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 119107
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-119107

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

RESULT 46

US-09-949-016-119143
Sequence 119143, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 119143
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-119143

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

RESULT 47

US-09-949-016-119179
Sequence 119179, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 119179
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-119179

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

RESULT 48

US-09-949-016-119215
Sequence 119215, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 119215
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-119215

Query Match 72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGAC 23
Db 289 AAAAAAAAAAAATTACTGTGAC 311

```
RESULT 49
US-09-949-016-142686/c
; Sequence 142686, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 142686
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-142686

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGACTGTGAC 23
        |||||
Db      90 AAAAAAAAAAGCCTGAAAGTGAC 68

RESULT 50
US-09-949-016-152429
; Sequence 152429, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152429
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-152429

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Search completed: December 14, 2005, 07:44:37
Job time : 66.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds
(without alignments)
800.703 Million cell updates/sec

Title: US-10-681-773-6

Perfect score: 25
Sequence: 1 aaaaaaaaaagcatgatctgacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Issued Parents NA: *
1: /cgn2_6/prodata/1/ina/1.COMB.seq: *
2: /cgn2_6/prodata/1/ina/5.COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/prodata/1/ina/H.COMB.seq: *
6: /cgn2_6/prodata/1/ina/PCBUS.COMB.seq: *
7: /cgn2_6/prodata/1/ina/PP.COMB.seq: *
8: /cgn2_6/prodata/1/ina/RE.COMB.seq: *
9: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	80.8	261	US-09-248-796A-11539	Sequence 11539, A
2	20.2	80.8	160552	US-09-593-828-11	Sequence 11, Appl
3	19.2	76.8	782	US-09-328-475C-144	Sequence 144, Appl
4	19.2	76.8	8050	US-09-491-362-11	Sequence 11, Appl
5	19.2	76.8	8050	US-09-874-562-11	Sequence 11, Appl
6	19.2	76.8	9997	US-09-949-016-12919	Sequence 12919, A
7	19.2	76.8	77994	US-09-949-016-12517	Sequence 12517, A
8	19.2	76.8	77994	US-09-949-016-16021	Sequence 16021, A
9	19.2	76.8	103894	US-09-949-016-14450	Sequence 14450, A
10	19.2	76.8	162465	US-09-949-016-14264	Sequence 14264, A
11	18.8	75.2	133	US-09-513-999C-21887	Sequence 21887, A
12	18.8	75.2	3246	US-09-005-180A-2	Sequence 21897, A
13	18.8	75.2	98439	US-09-949-016-13597	Sequence 13597, A
14	18.6	74.4	483	US-09-621-976-14600	Sequence 14600, A
15	18.6	74.4	601	US-09-949-016-43297	Sequence 43297, A
16	18.6	74.4	601	US-09-949-016-43298	Sequence 43298, A
17	18.6	74.4	601	US-09-949-016-43526	Sequence 43526, A
18	18.6	74.4	601	US-09-949-016-43527	Sequence 43527, A
19	18.6	74.4	601	US-09-949-016-43755	Sequence 43755, A
20	18.6	74.4	601	US-09-949-016-43756	Sequence 43756, A
21	18.6	74.4	28393	US-09-949-016-16980	Sequence 16980, A
22	18.6	74.4	40493	US-09-949-016-15453	Sequence 15453, A
23	18.6	74.4	60593	US-09-949-016-13779	Sequence 13779, A
24	18.6	74.4	91062	US-09-949-016-13019	Sequence 13019, A

25	18.6	74.4	106746	US-09-326-402C-1	Sequence 1, Appl
26	18.6	74.4	106746	US-09-326-402C-12	Sequence 12, Appl
27	18.6	74.4	116264	US-09-949-016-12756	Sequence 12756, A
28	18.6	74.4	136265	US-09-949-016-13001	Sequence 13001, A
29	18.6	74.4	147321	US-09-949-016-15450	Sequence 15450, A
30	18.6	74.4	157866	US-09-949-016-12982	Sequence 12982, A
31	18.6	74.4	157866	US-09-949-016-12983	Sequence 12983, A
32	18.6	74.4	157866	US-09-949-016-12984	Sequence 12984, A
33	18.6	74.4	300402	US-09-949-016-13632	Sequence 13632, A
34	18.6	74.4	314798	US-09-949-016-13539	Sequence 13539, A
35	18.4	73.6	80490	US-09-949-002-733	Sequence 733, App
36	18.4	73.6	82048	US-09-949-002-696	Sequence 696, App
37	18.4	73.6	82619	US-09-949-002-579	Sequence 579, App
38	18.4	73.6	112705	US-09-949-016-15630	Sequence 15630, A
39	18.4	73.6	146401	US-09-949-016-16151	Sequence 16151, A
40	18.2	72.8	601	US-09-949-016-94132	Sequence 94132, A
41	18.2	72.8	794	US-09-949-016-142866	Sequence 142866, A
42	18.2	72.8	794	US-09-328-475C-143	Sequence 143, App
43	18.2	72.8	1725	US-09-614-221A-266	Sequence 266, App
44	18.2	72.8	2914	US-08-454-097-11	Sequence 11, Appl
45	18.2	72.8	2914	US-08-185-359-11	Sequence 11, Appl
46	18.2	72.8	29970	US-09-949-016-14460	Sequence 14460, A
47	18.2	72.8	45571	US-09-949-016-16262	Sequence 16262, A
48	18.2	72.8	58356	US-09-949-016-15563	Sequence 15563, A
49	18.2	72.8	74804	US-09-949-016-15118	Sequence 15118, A
50	18.2	72.8	117937	US-09-949-016-12762	Sequence 12762, A
51	18.2	72.8	117937	US-09-949-016-15775	Sequence 15775, A
52	18.2	72.8	346112	US-09-949-016-13165	Sequence 13165, A
53	18.2	72.8	1830121	US-09-557-884-1	Sequence 1, Appl
54	18.2	72.8	1830121	US-09-643-990A-1	Sequence 1, Appl
55	18.2	72.8	1830121	US-10-158-865-1	Sequence 1, Appl
56	17.8	71.2	243	US-09-248-796A-8366	Sequence 8366, Ap
57	17.8	71.2	426	US-09-513-999C-34877	Sequence 34877, A
58	17.8	71.2	601	US-09-949-016-32593	Sequence 32593, A
59	17.8	71.2	601	US-09-949-016-155403	Sequence 155403, A
60	17.8	71.2	17607	US-09-949-016-15688	Sequence 15688, A
61	17.8	71.2	53336	US-09-949-016-12500	Sequence 12500, A
62	17.8	71.2	53337	US-09-949-016-16092	Sequence 16092, A
63	17.8	71.2	71574	US-09-949-016-15580	Sequence 15580, A
64	17.8	71.2	110000	US-09-830-902-1	Sequence 1, Appl
65	17.8	71.2	22750	US-09-949-016-17175	Sequence 17175, A
66	17.6	70.4	201	US-09-107-532A-439	Sequence 439, App
67	17.6	70.4	459	US-09-830-230A-8	Sequence 8, Appl
68	17.6	70.4	472	US-09-142-078-55	Sequence 55, Appl
69	17.6	70.4	472	US-09-357-141-55	Sequence 55, Appl
70	17.6	70.4	472	US-09-533-889-55	Sequence 55, Appl
71	17.6	70.4	472	US-09-142-080-55	Sequence 55, Appl
72	17.6	70.4	585	US-09-830-230A-7	Sequence 7, Appl
73	17.6	70.4	601	US-09-949-016-21880	Sequence 21880, A
74	17.6	70.4	601	US-09-949-016-26733	Sequence 26733, A
75	17.6	70.4	601	US-09-949-016-34398	Sequence 34398, A
76	17.6	70.4	601	US-09-949-016-46146	Sequence 46146, A
77	17.6	70.4	601	US-09-949-016-46147	Sequence 46147, A
78	17.6	70.4	601	US-09-949-016-46165	Sequence 46165, A
79	17.6	70.4	601	US-09-949-016-51161	Sequence 51161, A
80	17.6	70.4	601	US-09-949-016-64602	Sequence 64602, A
81	17.6	70.4	601	US-09-949-016-82241	Sequence 82241, A
82	17.6	70.4	601	US-09-949-016-83908	Sequence 83908, A
83	17.6	70.4	601	US-09-949-016-112284	Sequence 112284, A
84	17.6	70.4	601	US-09-949-016-135170	Sequence 135170, A
85	17.6	70.4	601	US-09-949-016-162170	Sequence 162170, A
86	17.6	70.4	601	US-09-949-016-162175	Sequence 162175, A
87	17.6	70.4	601	US-09-949-016-194175	Sequence 194175, A
88	17.6	70.4	601	US-09-949-016-204294	Sequence 204294, A
89	17.6	70.4	601	US-09-949-002-1489	Sequence 1489, Ap
90	17.6	70.4	601	US-09-949-002-3800	Sequence 3800, Ap
91	17.6	70.4	601	US-09-949-002-7691	Sequence 7691, Ap
92	17.6	70.4	601	US-09-949-002-9189	Sequence 9189, Ap
93	17.6	70.4	624	US-09-573-080A-224	Sequence 224, App
94	17.6	70.4	1139	US-08-755-587-40	Sequence 40, Appl
95	17.6	70.4	1404	US-09-614-221A-277	Sequence 277, App
96	17.6	70.4	1609	US-09-023-655-178	Sequence 178, App
97	17.6	70.4	2280	US-09-339-947A-2	Sequence 2, Appl

```

c 98 17.6 70.4 2476 3 US-09-774-639-31 Sequence 31, Appl
c 99 17.6 70.4 3001 3 US-09-539-33D-227 Sequence 227, App
c 100 17.6 70.4 4071 3 US-09-091-117-3 Sequence 3, Appl
c 101 17.6 70.4 4370 3 US-09-976-594-885 Sequence 885, App
c 102 17.6 70.4 4529 3 US-09-949-016-14004 Sequence 14004, A
c 103 17.6 70.4 5580 3 US-09-339-947A-3 Sequence 3, Appl
c 104 17.6 70.4 7195 3 US-09-949-016-12897 Sequence 12897, A
c 105 17.6 70.4 7198 3 US-09-949-016-15682 Sequence 15682, A
c 106 17.6 70.4 7505 3 US-09-078-294-13 Sequence 13, Appl
c 107 17.6 70.4 7699 3 US-09-949-016-2989 Sequence 2989, App
c 108 17.6 70.4 9226 3 US-09-949-016-12596 Sequence 12596, A
c 109 17.6 70.4 9226 3 US-09-949-016-16007 Sequence 16007, A
c 110 17.6 70.4 12779 3 US-09-949-016-13081 Sequence 13081, A
c 111 17.6 70.4 13006 3 US-09-949-016-14910 Sequence 14910, A
c 112 17.6 70.4 17132 3 US-09-949-016-15361 Sequence 15361, A
c 113 17.6 70.4 17930 3 US-09-949-016-11902 Sequence 11902, A
c 114 17.6 70.4 18351 3 US-09-949-016-16749 Sequence 16749, A
c 115 17.6 70.4 24150 3 US-09-949-016-12438 Sequence 12438, A
c 116 17.6 70.4 26843 3 US-09-949-016-17208 Sequence 17208, A
c 117 17.6 70.4 28802 3 US-09-949-016-14124 Sequence 14124, A
c 118 17.6 70.4 30868 3 US-09-949-016-13279 Sequence 13279, A
c 119 17.6 70.4 33624 3 US-09-949-016-15301 Sequence 15301, A
c 120 17.6 70.4 36907 3 US-09-949-016-12633 Sequence 12633, A
c 121 17.6 70.4 36913 3 US-09-949-016-15585 Sequence 15585, A
c 122 17.6 70.4 37215 3 US-09-949-016-15526 Sequence 15526, A
c 123 17.6 70.4 41965 3 US-09-949-016-13067 Sequence 13067, A
c 124 17.6 70.4 42376 3 US-09-949-016-16276 Sequence 16276, A
c 125 17.6 70.4 44244 3 US-09-949-016-11743 Sequence 11743, A
c 126 17.6 70.4 44245 3 US-09-949-016-13579 Sequence 13579, A
c 127 17.6 70.4 55068 3 US-09-949-002-691 Sequence 691, App
c 128 17.6 70.4 55068 3 US-09-949-002-778 Sequence 778, App
c 129 17.6 70.4 55195 3 US-09-949-016-15854 Sequence 15854, A
c 130 17.6 70.4 57932 3 US-09-949-002-619 Sequence 619, App
c 131 17.6 70.4 57933 3 US-09-949-002-822 Sequence 822, App
c 132 17.6 70.4 59828 3 US-09-949-016-16238 Sequence 16238, A
c 133 17.6 70.4 62987 3 US-09-949-016-17554 Sequence 17554, A
c 134 17.6 70.4 62987 3 US-09-949-016-12676 Sequence 12676, A
c 135 17.6 70.4 63714 3 US-09-949-016-16269 Sequence 16269, A
c 136 17.6 70.4 64321 3 US-09-949-016-16278 Sequence 16278, A
c 137 17.6 70.4 64321 3 US-09-949-016-13626 Sequence 13626, A
c 138 17.6 70.4 69763 3 US-09-949-016-12584 Sequence 12584, A
c 139 17.6 70.4 70000 3 US-09-851-896-3 Sequence 3, Appl
c 140 17.6 70.4 74940 3 US-09-949-016-17512 Sequence 17512, A
c 141 17.6 70.4 76390 3 US-09-949-016-16819 Sequence 16819, A
c 142 17.6 70.4 84558 3 US-09-949-016-15752 Sequence 15752, A
c 143 17.6 70.4 96327 3 US-09-949-016-16541 Sequence 16541, A
c 144 17.6 70.4 101894 3 US-09-949-016-12005 Sequence 12005, A
c 145 17.6 70.4 103894 3 US-09-949-016-14450 Sequence 14450, A
c 146 17.6 70.4 104520 3 US-09-949-016-13303 Sequence 13303, A
c 147 17.6 70.4 107140 3 US-09-949-016-14834 Sequence 14834, A
c 148 17.6 70.4 108310 3 US-09-949-016-16366 Sequence 16366, A
c 149 17.6 70.4 109925 3 US-09-949-016-13210 Sequence 13210, A
c 150 17.6 70.4 117410 3 US-09-949-016-12262 Sequence 12262, A
```

ALIGNMENTS

```

RESULT 1
; Sequence 11539, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
```

```

; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 11539
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-11539

Query Match      80.8%; Score 20.2; DB 3; Length 261;
Best Local Similarity 88.0%; Pred. No. 55;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
Db 133 AAAAAAAAAACATTAATGAGACAC 157

RESULT 2
; Sequence 11, Application US/09593828
; Patent No. 6852518
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; TITLE OF INVENTION: No. 6852518el Glycosyl Sulfotransferases
; FILE REFERENCE: 6510-138US1
; CURRENT APPLICATION NUMBER: US/09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 160552
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-593-828-11

Query Match      80.8%; Score 20.2; DB 3; Length 160552;
Best Local Similarity 88.0%; Pred. No. 81;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
Db 140431 AAAAAAAAAAGCCTTAATGAGACAC 140407

RESULT 3
; Sequence 144, Application US/09328475C
; Patent No. 6476207
; GENERAL INFORMATION:
; APPLICANT: Zhang, Jimmy
; APPLICANT: Astel, Jon H.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Endege, Wilson O.
; APPLICANT: Ford, Donna W.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; APPLICANT: Steinmann, Kathleen E.
; TITLE OF INVENTION: ARE DIFFERENTIALLY REGULATED IN PROSTATE CANCER
; FILE REFERENCE: 1532.002/200130.463
; CURRENT APPLICATION NUMBER: US/09/328,475C
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 144
; LENGTH: 782
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(782)
```

OTHER INFORMATION: n = A,T,C or G
US-09-328-475C-144

Query Match 76.8%; Score 19.2; DB 3; Length 782;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 469 AAAAAAAAAAGCATGATTGTGACA 492

RESULT 4
US-09-491-362-11/C
Sequence 11, Application US/09491362
Patent No. 6281017
GENERAL INFORMATION:
APPLICANT: Croceau, Rodney B
APPLICANT: Lange, Bernd M
TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
TITLE OF INVENTION: METHODS OF USE
FILE REFERENCE: WSUR14977
CURRENT APPLICATION NUMBER: US/09/491,362
CURRENT FILING DATE: 2000-01-26
EARLIER APPLICATION NUMBER: 60/118,349
EARLIER FILING DATE: 1999-02-03
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 8050
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-491-362-11

Query Match 76.8%; Score 19.2; DB 3; Length 8050;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 5603 AAAAAAAAAAGCATGATTGTGAAA 5580

RESULT 5
US-09-874-562-11/C
Sequence 11, Application US/09874562
Patent No. 6420159
GENERAL INFORMATION:
APPLICANT: Croceau, Rodney B
APPLICANT: Lange, Bernd M
TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
TITLE OF INVENTION: METHODS OF USE
FILE REFERENCE: WSUR17549
CURRENT APPLICATION NUMBER: US/09/874,562
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/491,362
PRIOR FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: 60/118,349
PRIOR FILING DATE: 1999-02-03
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 8050
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-874-562-11

Query Match 76.8%; Score 19.2; DB 3; Length 8050;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 5603 AAAAAAAAAAGCATGATTGTGAAA 5580

Db 5603 AAAAAAAAAAGCATGATTGTGAAA 5580

RESULT 6
US-09-949-016-12919
Sequence 12919, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12919
LENGTH: 9997
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(9997)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12919

Query Match 76.8%; Score 19.2; DB 3; Length 9997;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 2879 AAAAAAAAAAGCTTGTTGTGAGACA 2902

RESULT 7
US-09-949-016-12517/C
Sequence 12517, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12517
LENGTH: 77994
TYPE: DNA
ORGANISM: Human
US-09-949-016-12517

Query Match 76.8%; Score 19.2; DB 3; Length 77994;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 20751 AAAAAAAAAAGCATGATTGTGAAA 20728

RESULT 8
US-09-949-016-16021/C
; Sequence 16021, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16021
; LENGTH: 77994
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16021

Query Match 76.8%; Score 19.2; DB 3; Length 77994;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 20751 AAAAAAAAAAGATGATGTGAAA 20728

RESULT 9
US-09-949-016-14450
; Sequence 14450, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14450
; LENGTH: 103894
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(103894)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14450

Query Match 76.8%; Score 19.2; DB 3; Length 103894;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 30529 AAAAAAAAAAGTATGATGTGACA 30552

RESULT 10
US-09-949-016-14264/C
; Sequence 14264, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14264
; LENGTH: 162465
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14264

Query Match 76.8%; Score 19.2; DB 3; Length 162465;
Best Local Similarity 87.5%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGATTGTGACAC 25
DB 41779 AAAAAAAAAAGAAATTTGTGACAC 41756

RESULT 11
US-09-513-999C-21987/C
; Sequence 21987, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas, Mline Edwards, J.B.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; PATENT NO. 6783961
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21987
; LENGTH: 133
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-21987

Query Match 75.2%; Score 18.8; DB 3; Length 133;
Best Local Similarity 90.9%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGA 22
DB 119 AAAAAAAAAATCATATTGTGA 98

RESULT 12
US-09-005-180A-2/C
; Sequence 2, Application US/09005180A
; Patent No. 6124446
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

```

; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN VPS35/MEM3-RELATED PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,180A
; FILING DATE: Filed January 8, 1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0457 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3246 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGTUT08
; CLONE: 2641812
;
US-09-005-180A-2
Query Match 75.2%; Score 18.8; DB 3; Length 3246;
Best Local Similarity 90.9%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTGA 22
Db 3236 AAAAAAAAAATCATATTGTGA 3215

RESULT 13
US-09-949-016-13597
; Sequence 13597, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13597
; LENGTH: 98439
; TYPE: DNA
```

```

; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(98439)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13597

Query Match 75.2%; Score 18.8; DB 3; Length 98439;
Best Local Similarity 90.9%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTGA 22
Db 68111 AAAAAAAAAAGCTTGAATTGTA 68132

RESULT 14
US-09-621-976-14600/C
; Sequence 14600, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14600
; LENGTH: 483
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-14600

Query Match 74.4%; Score 18.6; DB 3; Length 483;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTGCAC 25
Db 419 AAAATATAATCATGTTGTGTAC 395

RESULT 15
US-09-949-016-43297/C
; Sequence 43297, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43297
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-43297

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db 429 AAAAAAAAAAGATTGTTACAC 405

RESULT 16
US-09-949-016-43298/C
; Sequence 43298, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43298
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-43298

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db 555 AAAAAAAAAAGATTGTTACAC 531

RESULT 17
US-09-949-016-43526/C
; Sequence 43526, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43526
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-43526

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db 429 AAAAAAAAAAGATTGTTACAC 405

RESULT 18
US-09-949-016-43527/C
; Sequence 43527, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43527
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-43527

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db 555 AAAAAAAAAAGATTGTTACAC 531

RESULT 19
US-09-949-016-43755/C
; Sequence 43755, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43755
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-43755

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db 429 AAAAAAAAAAGATTGTTACAC 405

RESULT 20
US-09-949-016-43756/C
; Sequence 43756, Application US/09949016
; Patent No. 6812339

```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43756
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-43756

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
Db 555 AAAAAAAAAAAGATTGTACAC 531

RESULT 21
US-09-949-016-16980
; Sequence 16980, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16980
; LENGTH: 28393
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(28393)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16980

Query Match          74.4%; Score 18.6; DB 3; Length 28393;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
Db 22875 AAAAAAAAAAACAATTGGAGCC 22899

RESULT 22
US-09-949-016-15453
; Sequence 15453, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15453
; LENGTH: 40493
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(40493)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15453

Query Match          74.4%; Score 18.6; DB 3; Length 40493;
Best Local Similarity 84.0%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
Db 13660 AAAAAAAAAAATCATGAAGTATAC 13684

RESULT 23
US-09-949-016-13779/C
; Sequence 13779, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13779
; LENGTH: 60593
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13779

Query Match          74.4%; Score 18.6; DB 3; Length 60593;
Best Local Similarity 84.0%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
Db 28450 AAAAAAAAAAAGATTGTGCAC 28426

RESULT 24
US-09-949-016-13019/C
; Sequence 13019, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
, TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
, TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
, FILE REFERENCE: CL001307
, CURRENT APPLICATION NUMBER: US/09/949,016
, CURRENT FILING DATE: 2000-04-14
, PRIOR APPLICATION NUMBER: 60/241,755
, PRIOR FILING DATE: 2000-10-20
, PRIOR APPLICATION NUMBER: 60/237,768
, PRIOR FILING DATE: 2000-10-03
, PRIOR APPLICATION NUMBER: 60/231,498
, PRIOR FILING DATE: 2000-09-08
, NUMBER OF SEQ ID NOS: 207012
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 13019
, LENGTH: 91062
, TYPE: DNA
, ORGANISM: Human
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (1)...(91062)
, OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13019

Query Match          74.4%; Score 18.6; DB 3; Length 91062;
Best Local Similarity 84.0%; Pred. NO. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  AAAAAAAAAAGCATGTTGTGACAC 25
Db      8568 ATAAAAAAAAATGCATGTTGTGACAC 8544

RESULT 25
US-09-326-402C-1
, Sequence 1, Application US/09326402C
, Patent No. 6759192
, GENERAL INFORMATION:
, APPLICANT: Blumenfeld, Marta
, APPLICANT: Bougueleret, Lydie
, APPLICANT: Chumakov, Ilya
, TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-1)
, FILE REFERENCE: GEN-T112XC1
, CURRENT APPLICATION NUMBER: US/09/326,402C
, CURRENT FILING DATE: 1999-06-04
, PRIOR APPLICATION NUMBER: 60/088,187
, PRIOR FILING DATE: 1998-06-05
, PRIOR APPLICATION NUMBER: 60/102,324
, PRIOR FILING DATE: 1998-09-28
, NUMBER OF SEQ ID NOS: 22
, SOFTWARE: Patent version 3.1
, SEQ ID NO 1
, LENGTH: 106746
, TYPE: DNA
, ORGANISM: Homo sapiens
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: 1..68647
, OTHER INFORMATION: 5'regulation region
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: 66647..68647
, OTHER INFORMATION: promoter
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: 97156..106746
, OTHER INFORMATION: 3'regulation region
, FEATURE:
, NAME/KEY: exon
, LOCATION: 68648..68741
, OTHER INFORMATION: exon0
, NAME/KEY: exon
, LOCATION: 70647..70794

, OTHER INFORMATION: exon1
, FEATURE:
, NAME/KEY: exon
, LOCATION: 82208..82296
, OTHER INFORMATION: exon2
, FEATURE:
, NAME/KEY: exon
, LOCATION: 83613..83823
, OTHER INFORMATION: exon3
, FEATURE:
, NAME/KEY: exon
, LOCATION: 85298..85417
, OTHER INFORMATION: exon4
, FEATURE:
, NAME/KEY: exon
, LOCATION: 86389..86445
, OTHER INFORMATION: exon5
, FEATURE:
, NAME/KEY: exon
, LOCATION: 87496..87522
, OTHER INFORMATION: exon6
, FEATURE:
, NAME/KEY: exon
, LOCATION: 87650..87775
, OTHER INFORMATION: exon6bis
, FEATURE:
, NAME/KEY: exon
, LOCATION: 88295..88383
, OTHER INFORMATION: exon7
, FEATURE:
, NAME/KEY: exon
, LOCATION: 89484..89649
, OTHER INFORMATION: exon8
, FEATURE:
, NAME/KEY: exon
, LOCATION: 92749..92755
, OTHER INFORMATION: exon9
, FEATURE:
, NAME/KEY: exon
, LOCATION: 92749..92883
, OTHER INFORMATION: exon9bis
, FEATURE:
, NAME/KEY: exon
, LOCATION: 95821..97155
, OTHER INFORMATION: exon9ter
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: 70647..70794
, OTHER INFORMATION: homology with genset EST : A241850
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: 68648..68741
, OTHER INFORMATION: homology with genset EST : A241850
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: 82208..82229
, OTHER INFORMATION: homology with genset EST : A241850
, NAME/KEY: allele
, LOCATION: 278
, OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
, FEATURE:
, NAME/KEY: allele
, LOCATION: 402
, OTHER INFORMATION: 99-1601-402 : polymorphic base A or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 472
, OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 2955
, OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
```



```
FEATURE:
NAME/KEY: allele
LOCATION: 12167
OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
FEATURE:
NAME/KEY: allele
LOCATION: 12536
OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
FEATURE:
NAME/KEY: allele
LOCATION: 17593
OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 17606
OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 22079
OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 28964
OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 29003
OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
FEATURE:
NAME/KEY: allele
LOCATION: 31077
OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 31766
OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 34791
OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 45751
OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 49847
OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 49855
OTHER INFORMATION: 99-1585-457 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 49886
OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
FEATURE:
NAME/KEY: allele
LOCATION: 49900
OTHER INFORMATION: 99-1585-412 : polymorphic base G or A
FEATURE:
NAME/KEY: allele
LOCATION: 49906
OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
FEATURE:
NAME/KEY: allele
LOCATION: 49921
OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
FEATURE:
NAME/KEY: allele
LOCATION: 49939
OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
FEATURE:
```

```
NAME/KEY: allele
LOCATION: 50256
OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
FEATURE:
NAME/KEY: allele
LOCATION: 54955
OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
FEATURE:
NAME/KEY: allele
LOCATION: 64239
OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 65436
OTHER INFORMATION: 99-1591-235 : polymorphic base A or G

Query Match 74.4% Score 18.6; DB 3; Length 106746;
Best Local Similarity 84.0% Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
Db 16222 AAAAAAAAAAGCATGATTATATACCC 16246

RESULT 26
US-09-326-402C-12
Sequence 12; Application US/09326402C
Patent No. 6759192
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marla
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-
CURRENT APPLICATION NUMBER: US/09/326,402C
CURRENT FILING DATE: 1999-06-04
PRIOR APPLICATION NUMBER: 60/088,187
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/102,324
PRIOR FILING DATE: 1998-09-28
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 106746
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..68647
OTHER INFORMATION: 5'regulation region
FEATURE:
NAME/KEY: misc feature
LOCATION: 66647..68647
OTHER INFORMATION: promoter
FEATURE:
NAME/KEY: misc feature
LOCATION: 97156..106746
OTHER INFORMATION: 3'regulation region
FEATURE:
NAME/KEY: exon
LOCATION: 68648..68741
OTHER INFORMATION: exon0
FEATURE:
NAME/KEY: exon
LOCATION: 70647..70794
OTHER INFORMATION: exon1
FEATURE:
NAME/KEY: exon
LOCATION: 82208..82296
OTHER INFORMATION: exon2
FEATURE:
NAME/KEY: exon
```

```
/ LOCATION: 83613..83823
/ OTHER INFORMATION: exon3
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 85298..85417
/ OTHER INFORMATION: exon4
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 86389..86445
/ OTHER INFORMATION: exon5
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 87496..87522
/ OTHER INFORMATION: exon6
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 87650..87775
/ OTHER INFORMATION: exon6bis
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 88295..88383
/ OTHER INFORMATION: exon7
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 89484..89649
/ OTHER INFORMATION: exon8
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 92749..97155
/ OTHER INFORMATION: exon9
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 92749..92883
/ OTHER INFORMATION: exon9bis
/ FEATURE:
/ NAME/KEY: exon
/ LOCATION: 95821..97155
/ OTHER INFORMATION: exon9ter
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 70647..70794
/ OTHER INFORMATION: homology with genset EST : A241850
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 68648..68741
/ OTHER INFORMATION: homology with genset EST : A241850
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 82208..82229
/ OTHER INFORMATION: homology with genset EST : A241850
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 278
/ OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 402
/ OTHER INFORMATION: 99-1601-402 : polymorphic base w= A or T
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 472
/ OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 2955
/ OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 12167
/ OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
/ NAME/KEY: allele
/ LOCATION: 12536

/ OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 17593
/ OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 17606
/ OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 22079
/ OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 28964
/ OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 29003
/ OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 31077
/ OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 31766
/ OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 34791
/ OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 45751
/ OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 49847
/ OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 49886
/ OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 49900
/ OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 49921
/ OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 49939
/ OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 50256
/ OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
/ NAME/KEY: allele
/ LOCATION: 54955
/ OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
```

```

FEATURE:
NAME/KEY: allele
LOCATION: 64239
OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 65436
OTHER INFORMATION: 99-1591-235 : polymorphic base A or G

Query Match      74.4%; Score 18.6; DB 3; Length 106746;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGCATGATTGTGACAC 25
Db      16222 AAAAAAAAAAGCATGATTATACCCC 16246

RESULT 27
US-09-949-016-12756
Sequence 12756, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001107
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12756
LENGTH: 136264
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1) .. (136264)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12756

Query Match      74.4%; Score 18.6; DB 3; Length 136264;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGCATGATTGTGACAC 25
Db      100609 AAAAAAAAAAGATTATGACAC 100633

RESULT 28
US-09-949-016-13001
Sequence 13001, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001107
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08

```

```

: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 13001
: LENGTH: 136265
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(136265)
: OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13001

Query Match          74.4%; Score 18.6; DB 3; Length 136265;
Best Local Similarity 84.0%; Pred. No.3,4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATGTGCAC 25
||||||| | ||| |||||
Db      100609 AAAAAAAAAAAGATGATGCAC 100633

RESULT 29
US-09-949-016-15450
: Sequence 15450, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: FILE REFERENCE: C1001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 15450
: LENGTH: 147321
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(147321)
: OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15450

Query Match          74.4%; Score 18.6; DB 3; Length 147321;
Best Local Similarity 84.0%; Pred. No.3,4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATGTGCAC 25
||||||| | ||| |||||
Db      54893 AAAAAAAAAATCATGCGTCTGCAC 54917

RESULT 30
US-09-949-016-12982
: Sequence 12982, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: FILE REFERENCE: C1001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768

```

```
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 12982
/ LENGTH: 157866
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-12982
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 157866;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGCAC 25
DB      107370 AAAAAAAAAAGATTGTTACAC 107394
```

```
RESULT 31
US-09-949-016-12983
/ Sequence 12983, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 12983
/ LENGTH: 157866
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-12983
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 157866;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGCAC 25
DB      107370 AAAAAAAAAAGATTGTTACAC 107394
```

```
RESULT 32
US-09-949-016-12984
/ Sequence 12984, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
```

```
/ SEQ ID NO: 12984
/ LENGTH: 157866
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-12984
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 157866;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGCAC 25
DB      107370 AAAAAAAAAAGATTGTTACAC 107394
```

```
RESULT 33
US-09-949-016-13632
/ Sequence 13632, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 13632
/ LENGTH: 300402
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)-(300402)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13632
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 300402;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGCAC 25
DB      154245 AAAAAAAAAAATATTGGACAC 154269
```

```
RESULT 34
US-09-949-016-13539/C
/ Sequence 13539, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 13539
```

```

; LENGTH: 314798
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(314798)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13539

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 3; Length 314798;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
Db 101212 AAGAAAGAAAGCATGATTCTGAGAC 101188

RESULT 35
US-09-949-002-733
; Sequence 733, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 733
; LENGTH: 80490
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(80490)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-733

Query Match
Best Local Similarity 73.6%; Score 18.4; DB 3; Length 80490;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGT 20
Db 71752 AAAAAAAAAAGTATGATTGT 71771

RESULT 36
US-09-949-002-696
; Sequence 696, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 696
; LENGTH: 82048
; TYPE: DNA
; ORGANISM: Human
```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(82048)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-696

Query Match
Best Local Similarity 73.6%; Score 18.4; DB 3; Length 82048;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGT 20
Db 71752 AAAAAAAAAAGTATGATTGT 71771

RESULT 37
US-09-949-002-579
; Sequence 579, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 579
; LENGTH: 82619
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(82619)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-002-579

Query Match
Best Local Similarity 73.6%; Score 18.4; DB 3; Length 82619;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGT 20
Db 71856 AAAAAAAAAAGTATGATTGT 71875

RESULT 38
US-09-949-016-15630
; Sequence 15630, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15630
; LENGTH: 112705
; TYPE: DNA
; ORGANISM: Human
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(112705)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15630
```

```
Query Match          73.6%; Score 18.4; DB 3; Length 112705;
Best Local Similarity 95.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      5 AAAAAAGCATGATTGTGACA 24
      |||||
Db      100098 AAAAAAGCATGATTGTGACA 100117
```

```
RESULT 39
US-09-949-016-16151
Sequence 16151, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16151
LENGTH: 146401
TYPE: DNA
ORGANISM: Human
US-09-949-016-16151
```

```
Query Match          73.6%; Score 18.4; DB 3; Length 146401;
Best Local Similarity 95.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAGCATGATTGT 20
      |||||
Db      14184 AAAAAAAGCATGATTGT 14203
```

```
RESULT 40
US-09-949-016-94132/c
Sequence 94132, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 94132
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-94132
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAGCATGATTGTGAC 23
      |||||
Db      466 AAAAAAAGCATGATTGAACTGAC 444
```

```
RESULT 41
US-09-949-016-142686/c
Sequence 142686, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 142686
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-142686
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAGCATGATTGTGAC 23
      |||||
Db      90 AAAAAAAGCCTGAAGTGTGAC 68
```

```
RESULT 42
US-09-328-475C-143/c
Sequence 143, Application US/09328475C
Patent No. 6476207
GENERAL INFORMATION:
APPLICANT: Zhang, Jimmy
APPLICANT: Astel, Jon H.
APPLICANT: Carroll III, Eddie
APPLICANT: Endege, Wilson O.
APPLICANT: Ford, Donna M.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
APPLICANT: Steimann, Kathleen E.
TITLE OF INVENTION: GENES AND GENE EXPRESSION PRODUCTS THAT
ARE DIFFERENTIALLY REGULATED IN PROSTATE CANCER
FILE REFERENCE: 1532.002/200130.463
CURRENT APPLICATION NUMBER: US/09/328,475C
PRIOR FILING DATE: 1999-06-09
NUMBER OF SEQ ID NOS: 341
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 143
LENGTH: 794
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(794)
OTHER INFORMATION: n = A,T,C or G
US-09-328-475C-143
```

Query Match 72.8%; Score 18.2; DB 3; Length 794;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGACA 24
DB 764 AAAAAAAAAAGCATGTTGACA 741

RESULT 43
US-09-614-221A-266
Sequence 266, Application US/09614221A
Patent No. 6723837
GENERAL INFORMATION:
APPLICANT: Karunanandaa, Balasubjini
APPLICANT: Yu, Jaehyuk
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
FILE REFERENCE: 16516.075
CURRENT APPLICATION NUMBER: US/09/614.221A
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/142,981
PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 626
SEQ ID NO 266
LENGTH: 1725
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-266

Query Match 72.8%; Score 18.2; DB 3; Length 1725;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGAC 23
DB 296 AAAAAAAAAAGCATGTTGAC 318

RESULT 44
US-08-454-097-11
Sequence 11, Application US/08454097
Patent No. 5686412
GENERAL INFORMATION:
APPLICANT: Hoekstra, Merl F.
TITLE OF INVENTION: Protein Kinases
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
City: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454.097
FILING DATE: 30-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/185,359
FILING DATE: 21-JAN-1994
APPLICATION NUMBER: US 08/008,001
FILING DATE: 21-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,783
FILING DATE: 03-JUL-1991

ATTORNEY/AGENT INFORMATION:
NAME: No. 5686412and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31853
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2914 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 265..1275
US-08-454-097-11

Query Match 72.8%; Score 18.2; DB 2; Length 2914;
Best Local Similarity 87.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGTTGACA 24
DB 1968 AAAAAAAAAAGCATGTTGACA 1990

RESULT 45
US-08-185-359-11
Sequence 11, Application US/08185359
Patent No. 6060296
GENERAL INFORMATION:
APPLICANT: Hoekstra, Merl F.
TITLE OF INVENTION: Protein Kinases
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
City: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,359
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,001
FILING DATE: 21-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,783
FILING DATE: 03-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 6060296and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31853
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2914 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 265..1275
US-08-185-359-11

Query Match          72.8%; Score 18.2; DB 3; Length 2914;
Best Local Similarity 87.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAAAAAAAGCATGATTGTGACA 24
Db      1968 AAAAAAAAAAGCATATTGTGCACA 1990

RESULT 46
US-09-949-016-14460/C
; Sequence 14460, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14460
; LENGTH: 29970
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(29970)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14460

Query Match          72.8%; Score 18.2; DB 3; Length 29970;
Best Local Similarity 87.0%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTGAC 23
Db      5401 AAAAAAAAAAGATGATTAAAGAC 5379

RESULT 47
US-09-949-016-16262
; Sequence 16262, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16262
```

```

; LENGTH: 45571
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16262

Query Match          72.8%; Score 18.2; DB 3; Length 45571;
Best Local Similarity 87.0%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTGAC 23
Db      30153 AATTAAAAAGCATATTGTGTAC 30175

RESULT 48
US-09-949-016-15563
; Sequence 15563, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15563
; LENGTH: 58356
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15563

Query Match          72.8%; Score 18.2; DB 3; Length 58356;
Best Local Similarity 87.0%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AAAAAAAAAAGCATGATTGTGACA 24
Db      56932 AAAAAAAAAAGCATATTGTGCACA 56954

RESULT 49
US-09-949-016-15118/C
; Sequence 15118, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15118
; LENGTH: 74804
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15118
```


Query Match 72.8%; Score 18.2; DB 3; Length 74804;
 Best Local Similarity 87.0%; Pred. No. 4.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTGAC 23
 DB 8346 AAAAAAAAAAGCATGAAAGAGAC 8324

RESULT 50
 US-09-949-016-12762/c
 ; Sequence 12762, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; FILE REFERENCE: CLO01307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; PRIOR FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 12762
 ; LENGTH: 117937
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-12762

Query Match 72.8%; Score 18.2; DB 3; Length 117937;
 Best Local Similarity 87.0%; Pred. No. 4.8e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGAC 23
 DB 18503 AAAAAAAAAAGCCTGAAGTGAC 18481

Search completed: December 14, 2005, 07:44:46
 Job time : 64.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds

(without alignments)
800.703 Million cell updates/sec

Title: US-10-681-773-7

Perfect score: 25
Sequence: 1 aaaaaaaaaagtcacattcagata 25Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1103057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/6H_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PTUTS_COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/PT_COMB.seq:*
- 8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
- 9: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	80.8	19299	3	US-09-949-016-13014
2	20.2	80.8	20229	3	US-09-949-016-16649
3	19.8	79.2	588	3	US-09-976-594-845
4	19.4	77.6	246240	2	US-08-724-394A-20
5	19.4	77.6	246240	2	US-08-724-394A-21
6	19.4	77.6	246240	2	US-08-724-394A-22
7	19.2	76.8	601	3	US-09-949-016-58666
8	19.2	76.8	601	3	US-09-949-016-58667
9	19.2	76.8	601	3	US-09-949-016-139296
10	19.2	76.8	601	3	US-09-949-016-139297
11	19.2	76.8	601	3	US-09-949-016-150576
12	19.2	76.8	601	3	US-09-949-016-150577
13	19.2	76.8	17629	3	US-09-949-016-16489
14	19.2	76.8	33663	3	US-09-949-016-15679
15	19.2	76.8	37254	3	US-09-949-016-15973
16	19.2	76.8	64137	3	US-09-949-016-14831
17	19.2	76.8	64171	3	US-09-949-016-12502
18	19.2	76.8	68702	3	US-09-949-016-16328
19	19.2	76.8	113100	3	US-09-949-016-12245
20	19.2	76.8	172677	3	US-09-949-016-13444
21	19.2	76.8	678533	3	US-09-949-016-14577
22	19.2	76.8	678533	3	US-09-949-016-14578
23	18.8	75.2	507	3	US-09-513-597A-18
24	18.8	75.2	601	3	US-09-949-016-33539

C 25	18.8	75.2	601	3	US-09-949-016-169310	Sequence 169330, A
C 26	18.8	75.2	601	3	US-09-949-016-169311	Sequence 169331, A
C 27	18.8	75.2	601	3	US-09-949-016-196111	Sequence 196111, A
C 28	18.8	75.2	51110	3	US-09-949-016-12069	Sequence 12069, A
C 29	18.8	75.2	51111	3	US-09-949-016-15724	Sequence 15724, A
C 30	18.8	75.2	173787	3	US-09-949-016-12542	Sequence 12542, A
C 31	18.8	75.2	173791	3	US-09-949-016-17302	Sequence 17302, A
C 32	18.8	75.2	173793	3	US-09-949-016-16513	Sequence 16513, A
C 33	18.6	74.4	313	3	US-09-270-767-10944	Sequence 30944, A
C 34	18.6	74.4	601	3	US-09-949-016-18398	Sequence 18398, A
C 35	18.6	74.4	601	3	US-09-949-016-18329	Sequence 91329, A
C 36	18.6	74.4	601	3	US-09-949-016-85296	Sequence 95296, A
C 37	18.6	74.4	601	3	US-09-949-016-95474	Sequence 95474, A
C 38	18.6	74.4	601	3	US-09-949-016-95652	Sequence 95652, A
C 39	18.6	74.4	601	3	US-09-949-016-95830	Sequence 95830, A
C 40	18.6	74.4	601	3	US-09-949-016-152568	Sequence 152568, A
C 41	18.6	74.4	601	3	US-09-949-016-163375	Sequence 163375, A
C 42	18.6	74.4	601	3	US-09-949-016-169887	Sequence 169887, A
C 43	18.6	74.4	601	3	US-09-949-016-183884	Sequence 183884, A
C 44	18.6	74.4	601	3	US-09-949-016-189301	Sequence 189301, A
C 45	18.6	74.4	601	3	US-09-949-016-191589	Sequence 191589, A
C 46	18.6	74.4	601	3	US-09-949-016-191767	Sequence 191767, A
C 47	18.6	74.4	601	3	US-09-949-016-191945	Sequence 191945, A
C 48	18.6	74.4	601	3	US-09-949-016-192123	Sequence 192123, A
C 49	18.6	74.4	601	3	US-09-949-016-192124	Sequence 192124, A
C 50	18.6	74.4	7766	3	US-09-949-016-11783	Sequence 11783, A
C 51	18.6	74.4	8190	3	US-09-949-016-14367	Sequence 14367, A
C 52	18.6	74.4	14551	3	US-09-949-016-12313	Sequence 12313, A
C 53	18.6	74.4	14551	3	US-09-949-016-16465	Sequence 16465, A
C 54	18.6	74.4	18891	3	US-09-949-016-16523	Sequence 16523, A
C 55	18.6	74.4	29350	3	US-09-949-016-11963	Sequence 11963, A
C 56	18.6	74.4	29350	3	US-09-949-016-11960	Sequence 11960, A
C 57	18.6	74.4	31390	3	US-09-949-016-15193	Sequence 15193, A
C 58	18.6	74.4	31399	3	US-09-949-016-16094	Sequence 16094, A
C 59	18.6	74.4	66524	3	US-09-949-001-32	Sequence 32, Appl
C 60	18.6	74.4	66525	3	US-09-949-001-38	Sequence 38, Appl
C 61	18.6	74.4	85854	3	US-09-949-016-12908	Sequence 12908, A
C 62	18.6	74.4	95109	3	US-09-949-016-16027	Sequence 16027, A
C 63	18.6	74.4	103447	3	US-09-949-016-16320	Sequence 16320, A
C 64	18.6	74.4	103447	3	US-09-949-016-14158	Sequence 14158, A
C 65	18.6	74.4	105210	3	US-09-949-016-17014	Sequence 17014, A
C 66	18.6	74.4	111424	3	US-09-949-016-17234	Sequence 17234, A
C 67	18.6	74.4	135010	3	US-09-949-016-17227	Sequence 17227, A
C 68	18.6	74.4	138693	3	US-09-949-016-16724	Sequence 16724, A
C 69	18.6	74.4	153866	3	US-09-949-016-14516	Sequence 14516, A
C 70	18.6	74.4	203475	3	US-09-949-016-14517	Sequence 14517, A
C 71	18.6	74.4	203475	3	US-09-949-016-14518	Sequence 14518, A
C 72	18.6	74.4	203475	3	US-09-949-016-14519	Sequence 14519, A
C 73	18.6	74.4	203475	3	US-09-949-016-17227	Sequence 17227, A
C 74	18.6	74.4	203475	3	US-09-949-016-17228	Sequence 17228, A
C 75	18.6	74.4	203475	3	US-09-949-016-17229	Sequence 17229, A
C 76	18.6	74.4	203475	3	US-09-949-016-13404	Sequence 13404, A
C 77	18.6	74.4	203475	3	US-09-949-016-13404	Sequence 13404, A
C 78	18.6	74.4	237863	3	US-09-949-016-17207	Sequence 17207, A
C 79	18.6	74.4	237863	3	US-09-949-016-12106	Sequence 12106, A
C 80	18.6	74.4	260286	3	US-09-949-016-13173	Sequence 13173, A
C 81	18.6	74.4	260293	3	US-09-949-016-13173	Sequence 13173, A
C 82	18.6	74.4	326791	3	US-09-949-016-13173	Sequence 13173, A
C 83	18.6	74.4	360470	3	US-09-513-999C-1801	Sequence 33538, A
C 84	18.4	73.6	103	3	US-09-949-016-196110	Sequence 196110, A
C 85	18.4	73.6	601	3	US-09-949-016-196110	Sequence 28, Appl
C 86	18.4	73.6	26664	3	US-09-554-805-28	Sequence 15057, A
C 87	18.4	73.6	26664	3	US-09-434-382-28	Sequence 15057, A
C 88	18.4	73.6	29558	3	US-09-248-796A-10567	Sequence 38286, A
C 89	18.4	73.6	29558	3	US-09-949-016-38286	Sequence 38287, A
C 90	18.2	72.8	234	3	US-09-949-016-38287	Sequence 38288, A
C 91	18.2	72.8	601	3	US-09-949-016-38288	Sequence 38289, A
C 92	18.2	72.8	601	3	US-09-949-016-38289	Sequence 38290, A
C 93	18.2	72.8	601	3	US-09-949-016-38290	Sequence 57819, A
C 94	18.2	72.8	601	3	US-09-949-016-38290	Sequence 149292, A
C 95	18.2	72.8	601	3	US-09-949-016-149292	
C 96	18.2	72.8	601	3	US-09-949-016-149292	
C 97	18.2	72.8	601	3	US-09-949-016-149292	

```
98 18.2 72.8 601 3 US-09-949-016-149293 Sequence 149293,
99 18.2 72.8 601 3 US-09-949-016-149294 Sequence 149294,
100 18.2 72.8 601 3 US-09-949-016-149295 Sequence 149295,
101 18.2 72.8 601 3 US-09-949-016-149296 Sequence 149296,
102 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
103 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
104 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
105 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
106 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
107 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
108 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
109 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
110 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
111 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
112 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
113 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
114 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
115 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
116 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
117 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
118 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
119 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
120 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
121 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
122 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
123 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
124 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
125 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
126 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
127 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
128 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
129 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
130 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
131 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
132 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
133 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
134 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
135 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
136 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
137 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
138 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
139 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
140 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
141 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
142 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
143 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
144 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
145 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
146 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
147 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
148 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
149 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
150 18.2 72.8 601 3 US-09-949-016-149312 Sequence 149312,
```

ALIGNMENTS

```
RESULT 1
US-09-949-016-13014
; Sequence 13014, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13014
; LENGTH: 19299
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13014
Query Match 80.8%; Score 20.2; DB 3; Length 19299;
Best Local Similarity 88.0%; Pred. No. 64;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Cy 1 AAAAAAAAAAATCCCAATTCAGATA 25
Db 19193 AAAAAAAAAAATCCCAATTCAGATA 19217
RESULT 2
US-09-949-016-16649
; Sequence 16649, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16649
; LENGTH: 20229
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16649
Query Match 80.8%; Score 20.2; DB 3; Length 20229;
Best Local Similarity 88.0%; Pred. No. 64;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Cy 1 AAAAAAAAAAATCCCAATTCAGATA 25
Db 19193 AAAAAAAAAAATCCCAATTCAGATA 19217
RESULT 3
US-09-976-594-845
; Sequence 845, Application US/0976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 845
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 231547.2
; US-09-976-594-845

Query Match          79.2% Score 19.8; DB 3; Length 588;
Best Local Similarity 91.3%; Pred. No. 80;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 23
Db 486 AAAAAAAAAAGTCCCAATTCA 508

RESULT 4
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereeto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Filts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H. CONTIG"
; US-08-724-394A-20

Query Match          77.6% Score 19.4; DB 2; Length 246240;
Best Local Similarity 95.2%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 21
Db 52401 AAAAAAAAAAGTCCCAATTCA 52421
```

```
RESULT 5
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereeto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Filts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H. CONTIG"
; US-08-724-394A-21

Query Match          77.6% Score 19.4; DB 2; Length 246240;
Best Local Similarity 95.2%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 21
Db 52401 AAAAAAAAAAGTCCCAATTCA 52421

RESULT 6
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
```

```

; TITLE OF INVENTION: Sequences and Antibodies Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Files, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.COMTIG"
;
US-08-724-394A-22

Query Match 77.6%; Score 19.4; DB 2; Length 246240;
Best Local Similarity 95.2%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCGA 21
Db 52401 AAAAAAAAAAGTCCCAATTCGA 52421

RESULT 7
US-09-949-016-58666/c
; Sequence 58666, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58666
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-58666

Query Match 76.8%; Score 19.2; DB 3; Length 601;
```

```

; TITLE OF INVENTION: Sequences and Antibodies Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Files, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.COMTIG"
;
US-08-724-394A-22

Query Match 77.6%; Score 19.4; DB 2; Length 246240;
Best Local Similarity 95.2%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCGA 21
Db 52401 AAAAAAAAAAGTCCCAATTCGA 52421

RESULT 7
US-09-949-016-58666/c
; Sequence 58666, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58666
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-58666

Query Match 76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 95.2%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCGA 21
Db 52401 AAAAAAAAAAGTCCCAATTCGA 52421

RESULT 7
US-09-949-016-58666/c
; Sequence 58666, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58667
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-58667

Query Match 76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 97.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 275 AAAAAAAAAAGTCCCAATTCAGAT 252

RESULT 8
US-09-949-016-58667/c
; Sequence 58667, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58667
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-58667

Query Match 76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 97.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 275 AAAAAAAAAAGTCCCAATTCAGATA 252

RESULT 8
US-09-949-016-139296
; Sequence 139296, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 139296
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 139296
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-139296

Query Match 76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 275 AAAAAAAAAAGTCCCAATTCAGATA 252
```

Db 148 AATAAAAAGTCCAAATTCGATA 171

```
RESULT 10
US-09-949-016-139297
; Sequence 139297, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 139297
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-139297
```

Query Match 76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAGTCCCAATTCGATA 25
Db 378 AATAAAAAGTCCAAATTCGATA 401

```
RESULT 11
US-09-949-016-150576/c
; Sequence 150576, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150576
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-150576
```

Query Match 76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAGTCCCAATTCGAT 24
Db 537 AAAAAAAGACCCATCTCAGAT 514

RESULT 12
US-09-949-016-150577/c

```
; Sequence 150577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150577
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-150577
```

Query Match 76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAGTCCCAATTCGAT 24
Db 407 AAAAAAAGACCCATCTCAGAT 384

```
RESULT 13
US-09-949-016-16489/c
; Sequence 16489, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16489
; LENGTH: 17629
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16489
```

Query Match 76.8%; Score 19.2; DB 3; Length 17629;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAGTCCCAATTCGATA 25
Db 8100 AAAAAAAGTCCCAATTCGATA 8077

```
RESULT 14
US-09-949-016-15679/c
; Sequence 15679, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15679
; LENGTH: 33663
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-15679

Query Match      76.8%; Score 19.2; DB 3; Length 33663;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 28040 AATAAAAAGTCCAAATTCGATA 28017

RESULT 15
US-09-949-016-15973/c
; Sequence 15973, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15973
; LENGTH: 37254
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(37254)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-15973

Query Match      76.8%; Score 19.2; DB 3; Length 37254;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 12100 AAAAAAAAAAGCCATCTCAGAT 12077

RESULT 16
US-09-949-016-14831/c
; Sequence 14831, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14831
; LENGTH: 64137
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(64137)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-14831

Query Match      76.8%; Score 19.2; DB 3; Length 64137;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 5280 AATAAAAAGTCTTAATTCAGATA 5257

RESULT 17
US-09-949-016-12502/c
; Sequence 12502, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12502
; LENGTH: 64171
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(64171)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-12502

Query Match      76.8%; Score 19.2; DB 3; Length 64171;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 5314 AATAAAAAGTCTTAATTCAGATA 5291

RESULT 18
US-09-949-016-16328/c
; Sequence 16328, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16328
; LENGTH: 64137
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(64137)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-16328

Query Match      76.8%; Score 19.2; DB 3; Length 64137;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 5280 AATAAAAAGTCTTAATTCAGATA 5257
```

```

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14831
; LENGTH: 64137
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(64137)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-14831

Query Match      76.8%; Score 19.2; DB 3; Length 64137;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 5280 AATAAAAAGTCTTAATTCAGATA 5257

RESULT 17
US-09-949-016-12502/c
; Sequence 12502, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12502
; LENGTH: 64171
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(64171)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-12502

Query Match      76.8%; Score 19.2; DB 3; Length 64171;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 5314 AATAAAAAGTCTTAATTCAGATA 5291

RESULT 18
US-09-949-016-16328/c
; Sequence 16328, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16328
; LENGTH: 64137
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(64137)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-16328

Query Match      76.8%; Score 19.2; DB 3; Length 64137;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 5280 AATAAAAAGTCTTAATTCAGATA 5257
```



```
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16328
; LENGTH: 68702
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16328

Query Match          76.8%; Score 19.2; DB 3; Length 68702;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 3495 AAAAAAAAAAGTCCCACTGCT 3472

RESULT 19
US-09-949-016-12245/C
; Sequence 12245, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12245
; LENGTH: 113100
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(113100)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12245

Query Match          76.8%; Score 19.2; DB 3; Length 113100;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13444
; LENGTH: 172677
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13444

Query Match          76.8%; Score 19.2; DB 3; Length 172677;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 72175 AAAAAAAAAATCCCAATCAAGAT 72152

RESULT 21
US-09-949-016-14577/C
; Sequence 14577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14577
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14577

Query Match          76.8%; Score 19.2; DB 3; Length 678533;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 567720 AAAAAAAAAAGTACCAATCAAT 567697

RESULT 22
US-09-949-016-14578/C
; Sequence 14578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14578
/ LENGTH: 678533
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(678533)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14578

Query Match          76.8%; Score 19.2; DB 3; Length 678533;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTTCAG 24
Db 567720 AAAAAAAAAAGTCCCAATTTCAG 567697

RESULT 23
US-09-513-597A-18/c
/ Sequence 18, Application US/09513597A
/ Patent No. 6770445
/ GENERAL INFORMATION:
/ APPLICANT: Scholler, Nathalie B.
/ APPLICANT: Helletrom, Ingegerd
/ APPLICANT: Helletrom, Karl Erik
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
/ TITLE OF INVENTION: CARCINOMAS
/ FILE REFERENCE: 730033,410
/ CURRENT APPLICATION NUMBER: US/09/513,597A
/ CURRENT FILING DATE: 2000-02-25
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 18
/ LENGTH: 507
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-513-597A-18

Query Match          75.2%; Score 18.8; DB 3; Length 507;
Best Local Similarity 90.9%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTTCAG 22
Db 266 AAAAAAAAAAGTCCCAATTTCAG 245

RESULT 24
US-09-949-016-33539
/ Sequence 33539, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
```

```
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 33539
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-33539

Query Match          75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTTCAG 22
Db 53 AAAAAAAAAAGTCCCAATTTCAG 74

RESULT 25
US-09-949-016-169330/c
/ Sequence 169330, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 169330
/ LENGTH: 601
/ TYPE: DNA
/ ORGANISM: Human
US-09-949-016-169330

Query Match          75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTTCAG 22
Db 407 AAAAAAAAAAGTCCCAATTTCAG 386

RESULT 26
US-09-949-016-169331/c
/ Sequence 169331, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 169331
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-169331
```

```
Query Match
Best Local Similarity 75.2%; Score 18.8; DB 3; Length 601;
Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAATCCCAATTGAG 22
Db 297 AAAAAAAAAATCCCAATTGAG 276
```

```
RESULT 27
US-09-949-016-196111
Sequence 196111, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 196111
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-196111
```

```
Query Match
Best Local Similarity 75.2%; Score 18.8; DB 3; Length 601;
Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTCCCAATTGAG 22
Db 53 AAAAAAAAAAGTCCCAATTGAG 74
```

```
RESULT 28
US-09-949-016-12069/C
Sequence 12069, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12069
LENGTH: 5110
TYPE: DNA
```

```
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(5110)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12069
```

```
Query Match
Best Local Similarity 75.2%; Score 18.8; DB 3; Length 5110;
Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 AAAAAAAAAATCCCAATTGAGAT 24
Db 11942 AAAAAAAAAATCCCAATTGAGAT 11921
```

```
RESULT 29
US-09-949-016-15724/C
Sequence 15724, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15724
LENGTH: 5111
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(5111)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15724
```

```
Query Match
Best Local Similarity 75.2%; Score 18.8; DB 3; Length 5111;
Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 3 AAAAAAAAAATCCCAATTGAGAT 24
Db 11942 AAAAAAAAAATCCCAATTGAGAT 11921
```

```
RESULT 30
US-09-949-016-12542/C
Sequence 12542, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
```

SEQ ID NO 12542
LENGTH: 173787
TYPE: DNA
ORGANISM: Human
US-09-949-016-12542

Query Match 75.2%; Score 18.8; DB 3; Length 173787;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAG 22
Db 170494 AAAAAAAAAAGTCCCAATTGAG 170473

RESULT 31
US-09-949-016-17302/c
Sequence 17302, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307

CURRENT FILING DATE: US/09/949,016

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 17302

LENGTH: 173791

TYPE: DNA

ORGANISM: Human

US-09-949-016-17302

Query Match 75.2%; Score 18.8; DB 3; Length 173791;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAG 22
Db 170494 AAAAAAAAAAGTCCCAATTGAG 170473

RESULT 32
US-09-949-016-16513/c
Sequence 16513, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: C1001307

CURRENT FILING DATE: US/09/949,016

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 16513

LENGTH: 177293

TYPE: DNA

ORGANISM: Human

US-09-949-016-16513

Query Match 75.2%; Score 18.8; DB 3; Length 177293;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAG 22
Db 71352 AAAAAAAAAAGTCCCAATTGAG 71331

RESULT 33

US-09-270-767-30944
Sequence 30944, Application US/09270767
Patent No. 6703491

GENERAL INFORMATION:

APPLICANT: Homburger et al.

TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

FILE REFERENCE: File Reference: 7326-094

CURRENT FILING DATE: US/09/270,767

NUMBER OF SEQ ID NOS: 62517

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 30944

LENGTH: 313

TYPE: DNA

ORGANISM: Drosophila melanogaster

US-09-270-767-30944

Query Match 74.4%; Score 18.6; DB 3; Length 313;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAGATA 25
Db 39 AAAAAAAAAAGTCCCAATTGAGATA 63

RESULT 34

US-09-949-016-18398
Sequence 18398, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: C1001307

CURRENT FILING DATE: US/09/949,016

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 18398

LENGTH: 601

TYPE: DNA

ORGANISM: Human

US-09-949-016-18398

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAGATA 25
Db 304 AAAAAAAAAAGTCCCAATTGAGATA 328

RESULT 35

US-09-949-016-91329

```
; Sequence 91329, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91329
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-91329

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATGCAATTCAGACA 326

RESULT 36
US-09-949-016-95296
; Sequence 95296, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95296
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-95296

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATGCAATTCAGACA 326

RESULT 37
US-09-949-016-95474
; Sequence 95474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95474
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-95474

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATGCAATTCAGACA 326

RESULT 38
US-09-949-016-95652
; Sequence 95652, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95652
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-95652

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATGCAATTCAGACA 326

RESULT 39
US-09-949-016-95830
; Sequence 95830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95830
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-95830

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATGCAATTCAGACA 326
```

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95830
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-95830

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATTGCAATTCAGACA 326

RESULT 40
US-09-949-016-152568
; Sequence 152568, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152568
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-152568

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 83 AAAAAAAAAAGACCTCAATTCAGACA 107

RESULT 41
US-09-949-016-163375/c
; Sequence 163375, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 163375
; LENGTH: 601
; TYPE: DNA
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 163375
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-163375

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 509 AAAAAAAAAATTGCAATTCAGAAA 485

RESULT 42
US-09-949-016-169887/c
; Sequence 169887, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169887
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-169887

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 398 AAAAAAAAAAGTCAATTCAGACA 374

RESULT 43
US-09-949-016-183884/c
; Sequence 183884, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 183884
; LENGTH: 601
; TYPE: DNA
```

ORGANISM: Human
US-09-949-016-183884

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 311 AAAAAAAAAAGCTTATTCAATA 287

RESULT 44
US-09-949-016-189301/c
Sequence 189301, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 189301
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-189301

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 479 AAAAAAAAAAGTACCAATGCTGTTA 455

RESULT 45
US-09-949-016-191589
Sequence 191589, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 191589
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-191589

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATTCATTCAGACA 326

RESULT 46
US-09-949-016-191767
Sequence 191767, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 191767
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-191767

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATTCATTCAGACA 326

RESULT 47
US-09-949-016-191945
Sequence 191945, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 191945
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-191945

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAATTCATTCAGACA 326

```

RESULT 48
US-09-949-016-192123
; Sequence 192123, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192123
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-192123

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 302 AAAAAAAAAAATTCGAATTCAGACA 326

RESULT 49
US-09-949-016-192724/c
; Sequence 192724, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192724
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-192724

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 31 AAAAAAAAAAATTCGAATTCAGATATA 7

RESULT 50
US-09-270-767-14715
; Sequence 14715, Application US/09270767

```

```

; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14715
; LENGTH: 995
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14715

Query Match          74.4%; Score 18.6; DB 3; Length 995;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 39 AAAAAAAAAAATTCATTCAGATA 63

Search completed: December 14, 2005, 07:44:55
Job time : 64.5 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 Search time 55.5 Seconds
(without alignments)
800.703 Million cell updates/sec

Title: US-10-681-773-8
Perfect score: 25
Sequence: 1 aaaaaaaaaagtcacatcagata 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/1_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfillseq.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	80.8	313	US-09-270-767-30944	Sequence 30944, A
2	20.2	80.8	601	US-09-949-016-95296	Sequence 95296, A
3	20.2	80.8	601	US-09-949-016-95474	Sequence 95474, A
4	20.2	80.8	601	US-09-949-016-95652	Sequence 95652, A
5	20.2	80.8	601	US-09-949-016-95830	Sequence 95830, A
6	20.2	80.8	601	US-09-949-016-191589	Sequence 191589, A
7	20.2	80.8	601	US-09-949-016-191767	Sequence 191767, A
8	20.2	80.8	601	US-09-949-016-191945	Sequence 191945, A
9	20.2	80.8	601	US-09-949-016-192123	Sequence 192123, A
10	20.2	80.8	995	US-09-270-767-14715	Sequence 14715, A
11	20.2	80.8	19299	US-09-949-016-13014	Sequence 13014, A
12	20.2	80.8	105210	US-09-949-016-14158	Sequence 14158, A
13	20.2	80.8	203475	US-09-949-016-14516	Sequence 14516, A
14	20.2	80.8	203475	US-09-949-016-14517	Sequence 14517, A
15	20.2	80.8	203475	US-09-949-016-14518	Sequence 14518, A
16	20.2	80.8	203475	US-09-949-016-14519	Sequence 14519, A
17	20.2	80.8	203475	US-09-949-016-17225	Sequence 17225, A
18	20.2	80.8	203475	US-09-949-016-17227	Sequence 17227, A
19	20.2	80.8	203475	US-09-949-016-17228	Sequence 17228, A
20	20.2	80.8	203475	US-09-949-016-17229	Sequence 17229, A
21	19.8	79.2	163662	US-09-949-016-12545	Sequence 12545, A
22	19.8	79.2	163662	US-09-949-016-13546	Sequence 13546, A
23	19.2	76.8	601	US-09-949-016-136203	Sequence 136203, A
24	19.2	76.8	601	US-09-949-016-136204	Sequence 136204, A

25	19.2	76.8	58356	3	US-09-949-016-15563	Sequence 15563, A
26	19.2	76.8	66788	3	US-09-949-001-37	Sequence 37, Appl
27	19.2	76.8	83938	3	US-09-949-016-16068	Sequence 16068, A
28	19.2	76.8	99797	3	US-09-949-016-15255	Sequence 15255, A
29	19.2	76.8	176006	3	US-09-949-016-16804	Sequence 16804, A
30	19.2	76.8	253375	3	US-09-949-016-12849	Sequence 12849, A
31	19.2	76.8	678533	3	US-09-949-016-14577	Sequence 14577, A
32	19.2	76.8	678533	3	US-09-949-016-14578	Sequence 14578, A
33	18.8	75.2	678533	3	US-09-949-016-18305	Sequence 48305, A
34	18.8	75.2	601	3	US-09-949-016-78602	Sequence 78602, A
35	18.8	75.2	601	3	US-09-949-016-78603	Sequence 78603, A
36	18.8	75.2	601	3	US-09-949-016-78604	Sequence 78604, A
37	18.8	75.2	601	3	US-09-949-016-78605	Sequence 78605, A
38	18.8	75.2	601	3	US-09-949-016-78606	Sequence 78606, A
39	18.8	75.2	346112	3	US-09-949-016-13165	Sequence 13165, A
40	18.8	75.2	784019	3	US-09-949-016-14033	Sequence 14033, A
41	18.8	75.2	828152	3	US-09-949-016-12777	Sequence 12777, A
42	18.6	74.4	102	3	US-08-956-171E-1391	Sequence 1391, Ap
43	18.6	74.4	102	3	US-08-781-986A-1391	Sequence 1391, Ap
44	18.6	74.4	123	3	US-09-475-947A-7	Sequence 7, Appl1
45	18.6	74.4	225	3	US-09-248-796A-10610	Sequence 10610, A
46	18.6	74.4	265	3	US-08-956-171E-1119	Sequence 1119, Ap
47	18.6	74.4	265	3	US-08-781-986A-1119	Sequence 1119, Ap
48	18.6	74.4	601	3	US-09-949-016-33145	Sequence 33145, A
49	18.6	74.4	601	3	US-09-949-016-15865	Sequence 65865, A
50	18.6	74.4	601	3	US-09-949-016-120859	Sequence 120859, A
51	18.6	74.4	601	3	US-09-949-016-131774	Sequence 131774, A
52	18.6	74.4	601	3	US-09-949-016-152522	Sequence 152522, A
53	18.6	74.4	601	3	US-09-949-016-169836	Sequence 169836, A
54	18.6	74.4	601	3	US-09-949-016-172864	Sequence 172864, A
55	18.6	74.4	601	3	US-09-949-016-189301	Sequence 189301, A
56	18.6	74.4	601	3	US-09-949-016-106677	Sequence 206677, A
57	18.6	74.4	2020	3	US-10-104-047-1171	Sequence 1171, Ap
58	18.6	74.4	6045	3	US-09-091-501B-7	Sequence 7, Appl1
59	18.6	74.4	10302	3	US-10-149-736-3	Sequence 23, Appl
60	18.6	74.4	10302	3	US-09-782-378A-23	Sequence 23, Appl
61	18.6	74.4	10320	3	US-09-091-501B-9	Sequence 9, Appl1
62	18.6	74.4	20229	3	US-09-949-016-16649	Sequence 16649, A
63	18.6	74.4	22120	3	US-09-949-016-16637	Sequence 16637, A
64	18.6	74.4	27430	3	US-09-949-016-13402	Sequence 13402, A
65	18.6	74.4	29350	3	US-09-949-016-11963	Sequence 11963, A
66	18.6	74.4	29350	3	US-09-949-016-17160	Sequence 17160, A
67	18.6	74.4	31096	3	US-08-956-171E-59	Sequence 59, Appl
68	18.6	74.4	31096	3	US-08-781-986A-59	Sequence 59, Appl
69	18.6	74.4	31390	3	US-09-949-016-15193	Sequence 15193, A
70	18.6	74.4	41804	3	US-09-949-016-12154	Sequence 12154, A
71	18.6	74.4	41893	3	US-09-949-016-14161	Sequence 14161, A
72	18.6	74.4	42250	3	US-09-949-016-17111	Sequence 17111, A
73	18.6	74.4	46718	3	US-09-816-093-3	Sequence 3, Appl1
74	18.6	74.4	47115	3	US-09-949-016-12278	Sequence 12278, A
75	18.6	74.4	47122	3	US-09-949-016-16502	Sequence 16502, A
76	18.6	74.4	50073	3	US-09-949-016-16026	Sequence 16026, A
77	18.6	74.4	64024	3	US-09-949-016-17553	Sequence 17593, A
78	18.6	74.4	66524	3	US-09-949-001-32	Sequence 32, Appl
79	18.6	74.4	66524	3	US-09-949-001-38	Sequence 38, Appl
80	18.6	74.4	93778	3	US-09-949-016-13195	Sequence 13195, A
81	18.6	74.4	95648	3	US-09-949-016-13195	Sequence 13195, A
82	18.6	74.4	131978	3	US-09-949-016-13117	Sequence 13117, A
83	18.6	74.4	133278	3	US-09-949-016-12554	Sequence 15254, A
84	18.6	74.4	147321	3	US-09-949-016-15450	Sequence 15450, A
85	18.6	74.4	163022	3	US-09-949-016-16515	Sequence 16515, A
86	18.6	74.4	235452	3	US-09-949-016-16515	Sequence 16515, A
87	18.6	74.4	422592	3	US-09-949-016-14182	Sequence 14182, A
88	18.4	73.6	594	3	US-09-370-838-262	Sequence 262, App
89	18.4	73.6	601	3	US-09-854-133-262	Sequence 262, App
90	18.4	73.6	601	3	US-09-949-016-44106	Sequence 44106, A
91	18.4	73.6	601	3	US-09-949-016-44135	Sequence 44135, A
92	18.4	73.6	601	3	US-09-949-016-44164	Sequence 44164, A
93	18.4	73.6	601	3	US-09-949-016-44193	Sequence 44193, A
94	18.4	73.6	601	3	US-09-949-016-44252	Sequence 44252, A
95	18.4	73.6	601	3	US-09-949-016-48306	Sequence 48306, A
96	18.4	73.6	1747	3	US-09-270-767-12074	Sequence 12074, A
97	18.4	73.6	22311	3	US-09-949-016-12344	Sequence 12344, A

```
98 18.4 73.6 22312 3 US-09-949-016-17217 Sequence 17217, A
C 99 18.4 73.6 36242 3 US-09-949-016-12996 Sequence 12996, A
C 100 18.4 73.6 36242 3 US-09-949-016-12997 Sequence 12997, A
C 101 18.4 73.6 36242 3 US-09-949-016-12998 Sequence 12998, A
C 102 18.4 73.6 36242 3 US-09-949-016-12999 Sequence 12999, A
C 103 18.4 73.6 36242 3 US-09-949-016-13000 Sequence 13000, A
C 104 18.4 73.6 105055 3 US-09-949-016-13001 Sequence 14000, A
C 105 18.2 72.8 588 3 US-09-976-594-845 Sequence 845, App
C 106 18.2 72.8 601 3 US-09-949-016-23520 Sequence 23520, A
C 107 18.2 72.8 601 3 US-09-949-016-15785 Sequence 35785, A
C 108 18.2 72.8 601 3 US-09-949-016-15785 Sequence 35785, A
C 109 18.2 72.8 601 3 US-09-949-016-35862 Sequence 35862, A
C 110 18.2 72.8 601 3 US-09-949-016-35863 Sequence 35863, A
C 111 18.2 72.8 601 3 US-09-949-016-35939 Sequence 35939, A
C 112 18.2 72.8 601 3 US-09-949-016-35940 Sequence 35940, A
C 113 18.2 72.8 601 3 US-09-949-016-36026 Sequence 36026, A
C 114 18.2 72.8 601 3 US-09-949-016-36027 Sequence 36027, A
C 115 18.2 72.8 601 3 US-09-949-016-162421 Sequence 162421, A
C 116 18.2 72.8 601 3 US-09-949-016-185962 Sequence 185962, A
C 117 18.2 72.8 601 3 US-09-949-016-197102 Sequence 197102, A
C 118 18.2 72.8 601 3 US-09-949-016-197103 Sequence 197103, A
C 119 18.2 72.8 601 3 US-09-949-016-197220 Sequence 197220, A
C 120 18.2 72.8 601 3 US-09-949-016-197221 Sequence 197221, A
C 121 18.2 72.8 601 3 US-09-949-016-197338 Sequence 197338, A
C 122 18.2 72.8 601 3 US-09-949-016-197339 Sequence 197339, A
C 123 18.2 72.8 601 3 US-09-949-016-197456 Sequence 197456, A
C 124 18.2 72.8 601 3 US-09-949-016-197457 Sequence 197457, A
C 125 18.2 72.8 3466 3 US-10-172-527A-11 Sequence 11, Appl
C 126 18.2 72.8 25321 3 US-09-949-016-12355 Sequence 12355, A
C 127 18.2 72.8 25321 3 US-09-949-016-14039 Sequence 14039, A
C 128 18.2 72.8 36699 3 US-09-684-960-8 Sequence 8, Appl1
C 129 18.2 72.8 58162 3 US-09-949-016-16289 Sequence 16289, A
C 130 18.2 72.8 84875 3 US-09-949-016-17334 Sequence 17334, A
C 131 18.2 72.8 84875 3 US-09-949-016-17335 Sequence 17335, A
C 132 18.2 72.8 84875 3 US-09-949-016-17336 Sequence 17336, A
C 133 18.2 72.8 85152 3 US-09-949-016-17337 Sequence 17337, A
C 134 18.2 72.8 85152 3 US-09-949-016-12665 Sequence 12665, A
C 135 18.2 72.8 85152 3 US-09-949-016-12666 Sequence 12666, A
C 136 18.2 72.8 85152 3 US-09-949-016-12667 Sequence 12667, A
C 137 18.2 72.8 85152 3 US-09-949-016-12668 Sequence 12668, A
C 138 18.2 72.8 88557 3 US-09-949-016-17028 Sequence 17028, A
C 139 18.2 72.8 133719 3 US-09-949-016-15092 Sequence 15092, A
C 140 18.2 72.8 331814 3 US-09-949-016-12008 Sequence 12008, A
C 141 18.2 72.8 331814 3 US-09-949-016-17056 Sequence 17056, A
C 142 17.8 71.2 282 3 US-09-117-257-65 Sequence 65, Appl
C 143 17.8 71.2 282 3 US-09-489-352-65 Sequence 65, Appl
C 144 17.8 71.2 467 3 US-09-536-059-20 Sequence 20, Appl
C 145 17.8 71.2 601 3 US-09-949-016-51765 Sequence 51765, A
C 146 17.8 71.2 601 3 US-09-949-016-51766 Sequence 51766, A
C 147 17.8 71.2 601 3 US-09-949-016-51767 Sequence 51767, A
C 148 17.8 71.2 601 3 US-09-949-016-55388 Sequence 55388, A
C 149 17.8 71.2 601 3 US-09-949-016-55389 Sequence 55389, A
C 150 17.8 71.2 601 3 US-09-949-016-173107 Sequence 173107, A
```

ALIGNMENTS

```
RESULT 1
US-09-270-767-30944
; Sequence 30944, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homurger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30944
; LENGTH: 313
; TYPE: DNA
```

```
; ORGANISM: Drosophila melanogaster
US-09-270-767-30944
```

```
Query Match 80.8%; Score 20.2; DB 3; Length 313;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTTCATTTCAGATA 25
Db 39 AAAAAAAAAACTTACATTTCAGATA 63
```

```
RESULT 2
US-09-949-016-95296
```

```
; Sequence 95296, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95296
; TYPE: DNA
```

```
US-09-949-016-95296
```

```
Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTTCATTTCAGATA 25
Db 302 AAAAAAAAAATTGCAATTCAGACA 326
```

```
RESULT 3
US-09-949-016-95474
```

```
; Sequence 95474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; PRIOR FILING DATE: 2000-04-14,755
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03,498
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95474
; LENGTH: 601
; TYPE: DNA
```

```
; ORGANISM: Human
US-09-949-016-95474
```

```
Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
```

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 302 AAAAAAAAAATTGCAATTCAGACA 326

RESULT 4

US-09-949-016-95652
; Sequence 95652, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95652
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-95652

Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 302 AAAAAAAAAATTGCAATTCAGACA 326

RESULT 5

US-09-949-016-95830
; Sequence 95830, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95830
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-95830

Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 302 AAAAAAAAAATTGCAATTCAGACA 326

RESULT 6

US-09-949-016-191589
; Sequence 191589, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 191589
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-191589

Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 302 AAAAAAAAAATTGCAATTCAGACA 326

RESULT 7

US-09-949-016-191767
; Sequence 191767, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 191767
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-191767

Query Match 80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 302 AAAAAAAAAATTGCAATTCAGACA 326

RESULT 8

US-09-949-016-191945
; Sequence 191945, Application US/09949016

```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 191945
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-191945

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 302 AAAAAAAAAAATTCATTCAGACA 326

RESULT 9
US-09-949-016-192123
; Sequence 192123, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192123
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-192123

Query Match      80.8%; Score 20.2; DB 3; Length 601;
Best Local Similarity 88.0%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 302 AAAAAAAAAAATTCATTCAGACA 326

RESULT 10
US-09-270-767-14715
; Sequence 14715, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
```

```
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14715
; LENGTH: 995
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14715

Query Match      80.8%; Score 20.2; DB 3; Length 995;
Best Local Similarity 88.0%; Pred. No. 90;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 39 AAAAAAAAAAATTCATTCAGATA 63

RESULT 11
US-09-949-016-13014
; Sequence 13014, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13014
; LENGTH: 19299
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13014

Query Match      80.8%; Score 20.2; DB 3; Length 19299;
Best Local Similarity 88.0%; Pred. No. 98;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 5165 AAAAAAAAAAATTCATTCAGATA 5189

RESULT 12
US-09-949-016-14158/c
; Sequence 14158, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14158
```

```

; LENGTH: 105210
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14158

Query Match      80.8%; Score 20.2; DB 3; Length 105210;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 94190 AAAAAAAAAAGTTCATTCAGATA 94166

RESULT 13
US-09-949-016-14516
; Sequence 14516, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14516
; LENGTH: 203475
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14516

Query Match      80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 113445 AAAAAAAAAAGTTCATTCAGACA 113469

RESULT 14
US-09-949-016-14517
; Sequence 14517, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14517
; LENGTH: 203475

; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14518

Query Match      80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 113445 AAAAAAAAAAGTTCATTCAGACA 113469

RESULT 15
US-09-949-016-14518
; Sequence 14518, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14518
; LENGTH: 203475
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14519

Query Match      80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 113445 AAAAAAAAAAGTTCATTCAGACA 113469

RESULT 16
US-09-949-016-14519
; Sequence 14519, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14519
; LENGTH: 203475
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14519

Query Match      80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 113445 AAAAAAAAAATTGCAATTCAGACA 113469

RESULT 17
US-09-949-016-17226
; Sequence 17226, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17226
; LENGTH: 203475
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17226

Query Match      80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 113445 AAAAAAAAAATTGCAATTCAGACA 113469

RESULT 18
US-09-949-016-17227
; Sequence 17227, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17227
; LENGTH: 203475
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17227

Query Match      80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 113445 AAAAAAAAAATTGCAATTCAGACA 113469

RESULT 19
US-09-949-016-17228
; Sequence 17228, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17228
; LENGTH: 203475
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17228

Query Match      80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 113445 AAAAAAAAAATTGCAATTCAGACA 113469

RESULT 20
US-09-949-016-17229
; Sequence 17229, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17229
; LENGTH: 203475
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(203475)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17229

Query Match          80.8%; Score 20.2; DB 3; Length 203475;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 113445 AAAAAAAAAAGTCCCAATTCAGACA 113469

RESULT 21
US-09-949-016-12545
; Sequence 12545, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12545
; LENGTH: 163662
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(163662)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12545

Query Match          79.2%; Score 19.8; DB 3; Length 163662;
Best Local Similarity 91.3%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 68588 AAAAAAAAAAGTCCCAATTCGA 68610

RESULT 22
US-09-949-016-13546
; Sequence 13546, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13546
; LENGTH: 163664
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(163664)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13546

Query Match          79.2%; Score 19.8; DB 3; Length 163664;
Best Local Similarity 91.3%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 68588 AAAAAAAAAAGTCCCAATTCGA 68610

RESULT 23
US-09-949-016-136203/C
; Sequence 136203, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136203
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(601)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-136203

Query Match          76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 90 AAAAAAAAAAGTCCCAATTCGAT 67

RESULT 24
US-09-949-016-136204/C
; Sequence 136204, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136204
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-136204

Query Match          76.8%; Score 19.2; DB 3; Length 601;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTTCAGAT 24
Db 147 AAAAAACAAGTCCAAATGCTGAT 124

RESULT 25
US-09-949-016-15563
; Sequence 15563, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15563
; LENGTH: 58356
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15563

Query Match          76.8%; Score 19.2; DB 3; Length 58356;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTTCAGAT 24
Db 5230 AAAAAACAAGTCCAAATGCTGAT 5253

RESULT 26
US-09-949-001-37/c
; Sequence 37, Application US/09949001
; Patent No. 6825336
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1000789
; CURRENT APPLICATION NUMBER: US/09/949,001
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/231,323
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 848
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 37
; LENGTH: 66788
; TYPE: DNA
; ORGANISM: Human
US-09-949-001-37

Query Match          76.8%; Score 19.2; DB 3; Length 66788;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTTCAGAT 24
Db 13501 AAAAAACAAGTCCATTTCAGAT 13478

RESULT 27
US-09-949-016-16068
; Sequence 16068, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16068
; LENGTH: 83938
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16068

Query Match          76.8%; Score 19.2; DB 3; Length 83938;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTTCAGAT 24
Db 47903 AAAAAACAAGTCCATTTCAGAT 47926

RESULT 28
US-09-949-016-15255/c
; Sequence 15255, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15255
; LENGTH: 99797
; TYPE: DNA
; ORGANISM: Human
FEATURE:
```



```

; NAME/KEY: misc_feature
; LOCATION: (1)...(99797)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15255

Query Match      76.8%; Score 19.2; DB 3; Length 99797;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 10934 AAAAAAAAAAGTCCCAATTAAGAT 10911

RESULT 29
US-09-949-016-16804
; Sequence 16804, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16804
; LENGTH: 176006
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176006)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16804

Query Match      76.8%; Score 19.2; DB 3; Length 176006;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 146925 AAAAAAAAAAGTCCCAAGTAAAT 146948

RESULT 30
US-09-949-016-12849
; Sequence 12849, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12849
; LENGTH: 253375
```

```

; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(253375)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12849

Query Match      76.8%; Score 19.2; DB 3; Length 253375;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 176294 AAAAAAAAAAGTCCCAAGTAAAT 176317

RESULT 31
US-09-949-016-14577/C
; Sequence 14577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14577
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14577

Query Match      76.8%; Score 19.2; DB 3; Length 678533;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 567720 AAAAAAAAAAGTACCAATTCAAAT 567697

RESULT 32
US-09-949-016-14578/C
; Sequence 14578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14578
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(678533)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14578

Query Match          76.8%; Score 19.2; DB 3; Length 678533;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCGAT 24
Db 567720 AAAAAAAAAAGTCCCAATTCGAT 567697

RESULT 33
US-09-949-016-48305
; Sequence 48305, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48305
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48305

Query Match          75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAG 22
Db 123 AAAAAAAAAAGTCCCAATTCAG 144

RESULT 34
US-09-949-016-78602
; Sequence 78602, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 78602
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-78602

Query Match          75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAG 22
Db 310 AAAAAAAAAAGTCCCAATTCAG 331

RESULT 35
US-09-949-016-78603
; Sequence 78603, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78603
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-78603

Query Match          75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAG 22
Db 177 AAAAAAAAAAGTCCCAATTCAG 198

RESULT 36
US-09-949-016-78604
; Sequence 78604, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78604
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-78604
```

Query Match 75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAG 22
|||||
Db 175 AAAAAAAAAAGTTCATTCAG 196

RESULT 37
US-09-949-016-78605
; Sequence 78605, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78605
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-78605

Query Match 75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAG 22
|||||
Db 173 AAAAAAAAAAGTTCATTCAG 194

RESULT 38
US-09-949-016-78606
; Sequence 78606, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78606
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-78606

Query Match 75.2%; Score 18.8; DB 3; Length 601;
Best Local Similarity 90.9%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAG 22
|||||
Db 172 AAAAAAAAAAGTTCATTCAG 193

RESULT 39
US-09-949-016-13165/C
; Sequence 13165, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13165
; LENGTH: 346112
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(346112)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13165

Query Match 75.2%; Score 18.8; DB 3; Length 346112;
Best Local Similarity 90.9%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAG 22
|||||
Db 294226 AAAAAAAAAAGTTCATTCAG 294205

RESULT 40
US-09-949-016-14033
; Sequence 14033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14033
; LENGTH: 784019
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(784019)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14033

Query Match 75.2%; Score 18.8; DB 3; Length 784019;

Best Local Similarity 90.9%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCG 22

Db 204051 AAAAAAAAAAGTTCATTCG 204072

RESULT 41

US-09-949-016-12777
Sequence 12777, Application US/09949016
Patent No. 681239
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03/231,498
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12777
LENGTH: 828152
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) ... (828152)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12777

Query Match 75.2%; Score 18.8; DB 3; Length 828152;
Best Local Similarity 90.9%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCG 22

Db 200184 AAAAAAAAAAGTTCATTCG 200205

RESULT 42

US-08-956-171E-1391
Sequence 1391, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunach
Gill H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 1391:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1391:
US-08-956-171E-1391

Query Match 74.4%; Score 18.6; DB 3; Length 102;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGATA 25

Db 32 AAAAAAATATTCATTCATATA 56

RESULT 43

US-08-781-986A-1391
Sequence 1391, Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 1391:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-1391

Query Match 74.4%; Score 18.6; DB 3; Length 102;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTGATA 25
Db 32 AAACAAATTAATTCATTGATA 56

RESULT 44

US-09-475-947A-7/c
; Sequence 7, Application US/09475947A
; Patent No. 6472154
; GENERAL INFORMATION:
; APPLICANT: Garner, Harold R.
; APPLICANT: Wren, Jonathan D.
; APPLICANT: Minna, John D.
; TITLE OF INVENTION: Polymorphic Repeats in Human Genes
; FILE REFERENCE: US/09/475,947A
; CURRENT APPLICATION NUMBER: US/09/475,947A
; CURRENT FILING DATE: 1999-12-31
; NUMBER OF SEQ ID NOS: 346
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 7
; LENGTH: 123
; TYPE: DNA
; ORGANISM: human
US-09-475-947A-7

Query Match 74.4%; Score 18.6; DB 3; Length 123;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTGATA 25
Db 78 AAACAAATTAATTCATTGATA 54

RESULT 45

US-09-248-796A-10610
; Sequence 10610, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO: 10610
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-10610

Query Match 74.4%; Score 18.6; DB 3; Length 225;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTGATA 25
Db 26 AAACAAATTAATTCATTGATA 50

RESULT 46

US-08-956-171E-1119/c
; Sequence 1119, Application US/08956171E

Patent No. 6593114

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 1119:
SEQUENCE CHARACTERISTICS:
LENGTH: 265 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1119:
US-08-956-171E-1119

Query Match 74.4%; Score 18.6; DB 3; Length 265;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTGATA 25
Db 131 AAACAAATTAATTCATTGATA 107

RESULT 47

US-08-781-986A-1119/c
; Sequence 1119, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

```

; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248BP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 1119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-781-986A-1119

Query Match          74.4%; Score 18.6; DB 3; Length 265;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 131 AAAAAAATAATTCATTGACATA 107

RESULT 48
US-09-949-016-33145
; Sequence 33145, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33145
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-33145

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 317 AAAAAAACACTTCTATTGACATA 341

RESULT 49
US-09-949-016-65865/c
; Sequence 65865, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```

```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65865
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-65865

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 271 AAAAAAATAATTCATTGACATA 247

RESULT 50
US-09-949-016-120859/c
; Sequence 120859, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 120859
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-120859

Query Match          74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTGACATA 25
Db 267 AAAAAAATAATTCATTGACATA 243

Search completed: December 14, 2005, 07:45:04
Job time : 64.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds

(without alignments)
800,703 Million cell updates/sec

Title: US-10-681-773-9

Perfect score: 25
Sequence: 1 aaaaaaaaaacattcatcatttaaac 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/1.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCpus.COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP.COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE.COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.8	79.2	601	3	US-09-949-016-163713 Sequence 163713, A
2	19.8	79.2	113538	3	US-09-949-016-16329 Sequence 16329, A
3	19.2	76.8	601	3	US-09-949-016-140314 Sequence 140314, A
4	19.2	76.8	38343	3	US-09-949-016-15714 Sequence 15714, A
5	19.2	76.8	87352	3	US-09-949-016-12053 Sequence 12053, A
6	19.2	76.8	87352	3	US-09-949-016-12721 Sequence 12721, A
7	19.2	76.8	87352	3	US-09-949-016-15692 Sequence 15692, A
8	19.2	76.8	87352	3	US-09-949-016-15693 Sequence 15693, A
9	19.2	76.8	129127	3	US-09-949-016-13461 Sequence 13461, A
10	18.6	74.4	204	3	US-09-248-796A-9761 Sequence 9761, A
11	18.6	74.4	444	3	US-09-543-681A-1457 Sequence 1457, A
12	18.6	74.4	2685	3	US-10-104-047-1143 Sequence 1143, A
13	18.6	74.4	47677	3	US-09-949-002-658 Sequence 658, A
14	18.6	74.4	50368	3	US-09-949-016-13256 Sequence 13256, A
15	18.6	74.4	81335	3	US-09-949-002-709 Sequence 709, A
16	18.6	74.4	83707	3	US-09-949-002-995 Sequence 995, A
17	18.6	74.4	151088	3	US-09-949-016-16240 Sequence 16240, A
18	18.6	74.4	194537	3	US-09-949-016-12928 Sequence 12928, A
19	18.6	74.4	201529	3	US-09-949-016-12740 Sequence 12740, A
20	18.6	74.4	235064	3	US-09-949-016-15390 Sequence 15390, A
21	18.6	74.4	298336	3	US-09-949-016-16600 Sequence 16600, A
22	18.6	74.4	1830121	3	US-09-557-884-1 Sequence 1, A
23	18.6	74.4	1830121	3	US-09-643-990A-1 Sequence 1, A
24	18.6	74.4	1830121	3	US-10-158-865-1 Sequence 1, A

25	18.4	73.6	63588	3	US-09-873-404-3 Sequence 3, Appl1
26	18.4	73.6	63588	3	US-10-243-735-3 Sequence 3, Appl1
27	18.2	72.8	601	3	US-09-949-016-144079 Sequence 144079, A
28	18.2	72.8	601	3	US-09-949-016-158012 Sequence 158012, A
29	18.2	72.8	601	3	US-09-949-016-158013 Sequence 158013, A
30	18.2	72.8	601	3	US-09-949-016-158017 Sequence 158017, A
31	18.2	72.8	601	3	US-09-949-002-1837 Sequence 3837, A
32	18.2	72.8	601	3	US-09-949-002-18470 Sequence 18470, A
33	18.2	72.8	673	3	US-09-270-767-1303 Sequence 7303, A
34	18.2	72.8	673	3	US-09-270-767-22585 Sequence 22585, A
35	18.2	72.8	1296	3	US-09-601-198-154 Sequence 154, A
36	18.2	72.8	3213	3	US-09-107-532A-62 Sequence 62, A
37	18.2	72.8	34172	3	US-09-949-016-14432 Sequence 14432, A
38	18.2	72.8	36620	3	US-09-949-016-16150 Sequence 16150, A
39	18.2	72.8	45716	3	US-08-965-048-5 Sequence 5, Appl1
40	18.2	72.8	45989	3	US-08-965-048-6 Sequence 6, Appl1
41	18.2	72.8	56694	3	US-09-949-016-12568 Sequence 12568, A
42	18.2	72.8	56702	3	US-09-949-016-15423 Sequence 15423, A
43	18.2	72.8	60141	3	US-09-949-016-15874 Sequence 15874, A
44	18.2	72.8	63467	3	US-09-949-002-693 Sequence 693, A
45	18.2	72.8	64518	3	US-09-949-016-17289 Sequence 17289, A
46	18.2	72.8	64518	3	US-09-949-002-843 Sequence 843, A
47	18.2	72.8	77036	3	US-09-949-016-16156 Sequence 16156, A
48	18.2	72.8	84839	3	US-09-949-016-15816 Sequence 15816, A
49	18.2	72.8	95318	3	US-09-949-016-11784 Sequence 11784, A
50	18.2	72.8	95318	3	US-09-949-016-13998 Sequence 13998, A
51	18.2	72.8	101558	3	US-09-949-016-12243 Sequence 12243, A
52	18.2	72.8	103934	3	US-09-949-016-14433 Sequence 14433, A
53	18.2	72.8	786431	3	US-09-751-389-3 Sequence 3, Appl1
54	17.8	71.2	298	3	US-09-513-399C-11549 Sequence 11549, A
55	17.8	71.2	601	3	US-09-949-016-18706 Sequence 28706, A
56	17.8	71.2	601	3	US-09-949-016-28707 Sequence 28707, A
57	17.8	71.2	601	3	US-09-949-016-17700 Sequence 77700, A
58	17.8	71.2	601	3	US-09-949-016-17701 Sequence 77701, A
59	17.8	71.2	601	3	US-09-949-002-1581 Sequence 1581, A
60	17.8	71.2	601	3	US-09-949-002-2371 Sequence 2371, A
61	17.8	71.2	601	3	US-09-949-002-3373 Sequence 3373, A
62	17.8	71.2	601	3	US-09-949-002-4154 Sequence 4154, A
63	17.8	71.2	601	3	US-09-949-002-4156 Sequence 4156, A
64	17.8	71.2	601	3	US-09-949-002-7015 Sequence 7015, A
65	17.8	71.2	10592	2	US-08-680-726A-51 Sequence 51, Appl1
66	17.8	71.2	10592	2	US-08-680-726A-52 Sequence 52, Appl1
67	17.8	71.2	10592	3	US-09-092-409-51 Sequence 51, Appl1
68	17.8	71.2	10592	3	US-09-092-409-52 Sequence 52, Appl1
69	17.8	71.2	15019	3	US-09-949-016-13571 Sequence 13571, A
70	17.8	71.2	37531	3	US-09-949-002-602 Sequence 602, A
71	17.8	71.2	37531	3	US-09-949-002-764 Sequence 764, A
72	17.8	71.2	40975	3	US-09-949-002-635 Sequence 635, A
73	17.8	71.2	40976	3	US-09-949-002-704 Sequence 704, A
74	17.8	71.2	76164	3	US-09-949-016-12288 Sequence 12288, A
75	17.8	71.2	76165	3	US-09-949-016-14005 Sequence 14005, A
76	17.8	71.2	10519	3	US-09-949-016-11769 Sequence 11769, A
77	17.8	71.2	119211	3	US-09-596-002-40 Sequence 40, Appl1
78	17.8	71.2	200663	3	US-09-949-016-12569 Sequence 12569, A
79	17.8	71.2	203093	3	US-09-949-016-14445 Sequence 14445, A
80	17.8	71.2	242487	3	US-09-949-002-687 Sequence 687, A
81	17.8	71.2	242488	3	US-09-949-016-17272 Sequence 17272, A
82	17.8	71.2	224288	3	US-09-949-002-841 Sequence 841, A
83	17.6	70.4	149	3	US-09-513-399C-21224 Sequence 21224, A
84	17.6	70.4	152	3	US-09-134-000C-2155 Sequence 2155, A
85	17.6	70.4	210	3	US-09-328-552-882 Sequence 882, A
86	17.6	70.4	216	3	US-09-248-796A-11250 Sequence 11250, A
87	17.6	70.4	220	3	US-09-513-399C-24541 Sequence 24541, A
88	17.6	70.4	222	3	US-09-248-796A-10723 Sequence 10723, A
89	17.6	70.4	243	3	US-09-543-681A-1659 Sequence 659, A
90	17.6	70.4	342	3	US-09-248-796A-13609 Sequence 13609, A
91	17.6	70.4	351	3	US-09-248-796A-10722 Sequence 10722, A
92	17.6	70.4	601	3	US-09-949-016-11962 Sequence 11962, A
93	17.6	70.4	601	3	US-09-949-016-13269 Sequence 13269, A
94	17.6	70.4	601	3	US-09-949-016-32505 Sequence 32505, A
95	17.6	70.4	601	3	US-09-949-016-32506 Sequence 32506, A
96	17.6	70.4	601	3	US-09-949-016-32507 Sequence 32507, A
97	17.6	70.4	601	3	US-09-949-016-34387 Sequence 34387, A

```
98 17.6 70.4 601 3 US-09-949-016-42463 Sequence 42463, A
99 17.6 70.4 601 3 US-09-949-016-50022 Sequence 50022, A
100 17.6 70.4 601 3 US-09-949-016-56189 Sequence 56189, A
101 17.6 70.4 601 3 US-09-949-016-65625 Sequence 65625, A
102 17.6 70.4 601 3 US-09-949-016-89365 Sequence 89365, A
103 17.6 70.4 601 3 US-09-949-016-134674 Sequence 134674, A
104 17.6 70.4 601 3 US-09-949-016-134675 Sequence 134675, A
105 17.6 70.4 601 3 US-09-949-016-140450 Sequence 140450, A
106 17.6 70.4 601 3 US-09-949-016-140451 Sequence 140451, A
107 17.6 70.4 601 3 US-09-949-016-145340 Sequence 145340, A
108 17.6 70.4 601 3 US-09-949-016-176820 Sequence 176820, A
109 17.6 70.4 601 3 US-09-949-016-176821 Sequence 176821, A
110 17.6 70.4 601 3 US-09-949-016-187426 Sequence 187426, A
111 17.6 70.4 601 3 US-09-949-016-187427 Sequence 187427, A
112 17.6 70.4 601 3 US-09-949-016-187428 Sequence 187428, A
113 17.6 70.4 601 3 US-09-949-016-190074 Sequence 190074, A
114 17.6 70.4 601 3 US-09-949-016-191226 Sequence 191226, A
115 17.6 70.4 601 3 US-09-949-001-119 Sequence 119, App
116 17.6 70.4 601 3 US-09-949-001-575 Sequence 575, App
117 17.6 70.4 660 3 US-09-328-111-784 Sequence 784, App
118 17.6 70.4 675 3 US-09-543-681A-1834 Sequence 1834, App
119 17.6 70.4 700 3 US-09-735-271-1159 Sequence 1159, App
120 17.6 70.4 750 3 US-09-270-767-8130 Sequence 8130, App
121 17.6 70.4 750 3 US-09-270-767-23412 Sequence 23412, A
122 17.6 70.4 955 3 US-08-936-165A-149 Sequence 149, App
123 17.6 70.4 1001 3 US-08-641-638-78 Sequence 78, App
124 17.6 70.4 1001 3 US-10-170-097-78 Sequence 78, App
125 17.6 70.4 1477 3 US-09-620-312D-831 Sequence 831, App
126 17.6 70.4 1572 3 US-09-907-794A-340 Sequence 340, App
127 17.6 70.4 1572 3 US-09-905-125A-340 Sequence 340, App
128 17.6 70.4 1572 3 US-09-902-775A-340 Sequence 340, App
129 17.6 70.4 1572 3 US-09-906-618-340 Sequence 340, App
130 17.6 70.4 1572 3 US-09-903-603A-340 Sequence 340, App
131 17.6 70.4 1572 3 US-09-904-603A-340 Sequence 340, App
132 17.6 70.4 1572 3 US-09-904-920A-340 Sequence 340, App
133 17.6 70.4 1572 3 US-09-905-281A-340 Sequence 340, App
134 17.6 70.4 1572 3 US-09-906-618-340 Sequence 340, App
135 17.6 70.4 1572 3 US-09-906-646-340 Sequence 340, App
136 17.6 70.4 1572 3 US-09-904-462-340 Sequence 340, App
137 17.6 70.4 1572 3 US-09-902-736A-340 Sequence 340, App
138 17.6 70.4 1572 3 US-09-906-722A-340 Sequence 340, App
139 17.6 70.4 1618 2 US-08-832-883-50 Sequence 50, App
140 17.6 70.4 1618 2 US-08-832-877-50 Sequence 50, App
141 17.6 70.4 1882 2 US-08-257-073-12 Sequence 12, App
142 17.6 70.4 1883 3 US-09-398-179-3 Sequence 3, App
143 17.6 70.4 1883 3 US-09-398-179-4 Sequence 4, App
144 17.6 70.4 1884 2 US-08-257-073-8 Sequence 8, App
145 17.6 70.4 2064 3 US-09-248-796A-8923 Sequence 8923, App
146 17.6 70.4 2238 3 US-09-710-279-1741 Sequence 1741, App
147 17.6 70.4 2397 3 US-09-134-001C-2080 Sequence 2080, App
148 17.6 70.4 2435 3 US-09-976-594-286 Sequence 286, App
149 17.6 70.4 2435 3 US-09-919-039-142 Sequence 142, App
150 17.6 70.4 2455 3 US-09-103-429A-1 Sequence 1, App
```

ALIGNMENTS

```
RESULT 1
US-09-949-016-163713
; Sequence 163713, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 163713
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-163713

Query Match 79.2%; Score 19.8; DB 3; Length 601;
Best Local Similarity 91.3%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTCATCATTTAA 23
Db 311 AAAAAAAAACTCATTAATTTAA 333
```

```
RESULT 2
US-09-949-016-16329/c
; Sequence 16329, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16329
; LENGTH: 113538
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc.feature
; LOCATION: (1)...(113538)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16329
```

```
Query Match 79.2%; Score 19.8; DB 3; Length 113538;
Best Local Similarity 91.3%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTCATCATTTAA 23
Db 84584 AAAAAAAAACTCATTAATTTAA 84562
```

```
RESULT 3
US-09-949-016-140314/c
; Sequence 140314, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```



```

RESULT 7
US-09-949-016-15692
; Sequence 15692, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231, 498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 15692
; LENGTH: 87352
; TYPE: DNA

```

ORGANISM: Human
US-09-949-016-15692

Query Match 76.8%; Score 19.2; DB 3; Length 87352;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
Db 54330 AAAAAAAAAAGTGCATCATTTAA 54353

RESULT 8

US-09-949-016-15693
Sequence 15693, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CU001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15693
LENGTH: 87352
TYPE: DNA
ORGANISM: Human
US-09-949-016-15693

Query Match 76.8%; Score 19.2; DB 3; Length 87352;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
Db 54330 AAAAAAAAAAGTGCATCATTTAA 54353

RESULT 9

US-09-949-016-13481/C
Sequence 13481, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CU001307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13481
LENGTH: 129127
TYPE: DNA
ORGANISM: Human
US-09-949-016-13481

Query Match 76.8%; Score 19.2; DB 3; Length 129127;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTTCATCATTTAAAC 25
Db 54099 AAAAAAAAACTTAAGTATTTAAAC 54076

RESULT 10

US-09-248-796A-9761/C
Sequence 9761, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 9761
LENGTH: 204
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-9761

Query Match 74.4%; Score 18.6; DB 3; Length 204;
Best Local Similarity 84.0%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 140 AAAAAAAAAATTCTTACCATTTAAAC 116

RESULT 11

US-09-543-681A-1457
Sequence 1457, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 1457
LENGTH: 444
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-1457

Query Match 74.4%; Score 18.6; DB 3; Length 444;
Best Local Similarity 84.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 83 AAAAAATACCTCCTCATCATTTAAAC 107

RESULT 12

US-10-104-047-1143/C
Sequence 1143, Application US/10104047
Patent No. 6943241
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE

```

; TITLE OF INVENTION: No. 6943241e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1143
; LENGTH: 2685
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1143

Query Match      74.4%; Score 18.6; DB 3; Length 2685;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTTCATCATTTAAAC 25
Db      1589 AAAAAAAAACTTCATCATTTAAAC 1565

RESULT 13
US-09-949-002-668
; Sequence 668, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 668
; LENGTH: 47677
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-668

Query Match      74.4%; Score 18.6; DB 3; Length 47677;
Best Local Similarity 84.0%; Pred. No. 4.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTTCATCATTTAAAC 25
Db      21444 AAAAAAAAACTTCATCATTTAAAC 21468

RESULT 14
US-09-949-016-13256/c
; Sequence 13256, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 13256
; LENGTH: 50368
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13256

Query Match      74.4%; Score 18.6; DB 3; Length 50368;
Best Local Similarity 84.0%; Pred. No. 4.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTTCATCATTTAAAC 25
Db      31763 AAAAAAAAACTTCATCATTTAAAC 31739

RESULT 15
US-09-949-002-709/c
; Sequence 709, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 709
; LENGTH: 81335
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-709

Query Match      74.4%; Score 18.6; DB 3; Length 81335;
Best Local Similarity 84.0%; Pred. No. 4.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTTCATCATTTAAAC 25
Db      23936 AAAAAAAAACTTCATCATTTAAAC 23912

RESULT 16
US-09-949-002-595/c
; Sequence 595, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 595
; LENGTH: 83707
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-595

Query Match      74.4%; Score 18.6; DB 3; Length 83707;
Best Local Similarity 84.0%; Pred. No. 4.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTTCATCATTTAAAC 25
```

Db 25647 AAAAAAAAACTATATATTTAAAC 25623

```
RESULT 17
US-09-949-016-16240/C
; Sequence 16240, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16240
; LENGTH: 151088
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(151088)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16240
```

Query Match 74.4%; Score 18.6; DB 3; Length 151088;
Best Local Similarity 84.0%; Pred. No. 4.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTCATCTTTAAAC 25
Db 126129 AAAAAAGACTCACTATTAAC 126105

```
RESULT 18
US-09-949-016-12928/C
; Sequence 12928, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12928
; LENGTH: 194537
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(194537)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12928
```

Query Match 74.4%; Score 18.6; DB 3; Length 194537;
Best Local Similarity 84.0%; Pred. No. 4.6e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTCATCTTTAAAC 25
Db 82553 AAAAAAGACTCACTATTTAAAC 82529

```
RESULT 19
US-09-949-016-12740/C
; Sequence 12740, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12740
; LENGTH: 201529
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(201529)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12740
```

Query Match 74.4%; Score 18.6; DB 3; Length 201529;
Best Local Similarity 84.0%; Pred. No. 4.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTCATCTTTAAAC 25
Db 89553 AAAAAAGACTCACTATTTAAAC 89529

```
RESULT 20
US-09-949-016-15390/C
; Sequence 15390, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15390
; LENGTH: 235064
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(235064)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15390
```

Query Match 74.4%; Score 18.6; DB 3; Length 235064;
Best Local Similarity 84.0%; Pred. No. 4.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAAC 25
DB 222405 AAAAAACACCTTCATTAATTAAAC 222381

RESULT 21
US-09-949-016-16600
Sequence 16600, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C0001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/223,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 16600
LENGTH: 298336
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(298336)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16600

Query Match 74.4%; Score 18.6; DB 3; Length 298336;
Best Local Similarity 84.0%; Pred. No. 4.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAAC 25
DB 35459 AAAAAACACCTTCATCATCTCAGC 35483

RESULT 22
US-09-557-884-1/c
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Pleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 74.4%; Score 18.6; DB 3; Length 1830121;
Best Local Similarity 84.0%; Pred. No. 4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAAC 25
DB 1529389 AAAAAACACCTTCAGCATTTAAAC 1529365

RESULT 23
US-09-643-990A-1/c
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Pleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```

;
;   LENGTH: 1830121 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;
;   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match      74.4%; Score 18.6; DB 3; Length 1830121;
Best Local Similarity 84.0%; Pred. No. 4e+02;
Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db      1529389 AAGAAAAAACTCAGCATTTAAAC 1529365

RESULT 24
US-10-158-865-1/c
; Sequence 1, Application US/10158865
; Patent No. 6846651
; GENERAL INFORMATION:
;   APPLICANT: Fleischmann et al.
;   TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
;   Patent No. 6846651
;   TITLE OF INVENTION: Thereof, and Uses Thereof
;   FILE REFERENCE: PB186P2CID1
;   CURRENT APPLICATION NUMBER: US/10/158, 865
;   CURRENT FILING DATE: 2002-06-03
;   PRIOR APPLICATION NUMBER: US 09/557,884
;   PRIOR FILING DATE: 2000-04-25
;   PRIOR APPLICATION NUMBER: US 08/476,102
;   PRIOR FILING DATE: 1995-06-07
;   PRIOR APPLICATION NUMBER: US 08/426,787
;   PRIOR FILING DATE: 1995-04-21
;   NUMBER OF SEQ ID NOS: 1
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus Influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9921)..(9921)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10150)..(10150)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (29298)..(29298)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36543)..(36543)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36551)..(36551)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36636)..(36636)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (40808)..(40810)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:

;
;   NAME/KEY: misc_feature
;   LOCATION: (44416)..(44416)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44905)..(44905)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44975)..(44975)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45593)..(45593)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (47036)..(47036)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45732)..(45732)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51334)..(51334)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51602)..(51602)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51786)..(51786)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51805)..(51805)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (55369)..(55369)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65309)..(65309)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65313)..(65313)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (80024)..(80024)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10091)..(10091)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (102696)..(102696)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (105121)..(105121)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (107248)..(107248)
;   OTHER INFORMATION: n equals a,t,c, or g
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
```

```
Query Match 74.4%; Score 18.6; DB 3; Length 1830121;
Best Local Similarity 84.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 1529389 AAGAAAAAATTCGATTAAC 1529365
```

```
RESULT 25
US-09-873-404-3
; Sequence 3, Application US/09873404
; Patent No. 650656
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CU001212-CIP
; CURRENT APPLICATION NUMBER: US/09/873,404
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 63588
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(63588)
; OTHER INFORMATION: n = A,T,C or G
US-09-873-404-3
```

```
Query Match 73.6%; Score 18.4; DB 3; Length 63588;
Best Local Similarity 95.0%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTTCATCATTT 20
Db 60852 AAAAAAAAACTTCATCAGT 60871
```

```
RESULT 26
US-10-243-735-3
; Sequence 3, Application US/10243735
; Patent No. 6706510
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CU001212DIV
; CURRENT APPLICATION NUMBER: US/10/243,735
; CURRENT FILING DATE: 2002-09-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 63588
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(63588)
; OTHER INFORMATION: n = A,T,C or G
US-10-243-735-3
```

```
Query Match 73.6%; Score 18.4; DB 3; Length 63588;
Best Local Similarity 95.0%; Pred. No. 5.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTTCATCATTT 20
Db 60852 AAAAAAAAACTTCATCAGT 60871
```

```
RESULT 27
US-09-949-016-144079
; Sequence 144079, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 144079
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-144079

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 10 AAAAAAAAAATCTTACTTCATTTAA 32

RESULT 28
US-09-949-016-158012/c
; Sequence 158012, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 158012
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-158012

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 390 AAAAAAAAAATCTTACTTCATTTAA 368

RESULT 29
US-09-949-016-158013/c
; Sequence 158013, Application US/09949016
```

```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 158013
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-158013

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 51 AAAAAAAAAATCTTACTTCATTTAA 29

RESULT 30
US-09-949-016-195372
; Sequence 195372, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 195372
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-195372

Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 498 AAAAAAAAAATCTTACTTCATTTAA 520

RESULT 31
US-09-949-002-3837
; Sequence 3837, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
```



```
;; TITLE OF INVENTION: AND USES THEREOF
;; FILE REFERENCE: CL000790
;; CURRENT APPLICATION NUMBER: US/09/949,002
;; CURRENT FILING DATE: 2000-01-28
;; PRIOR APPLICATION NUMBER: 60/231,401
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 10823
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3837.
;; LENGTH: 601
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-002-3837
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTTCATCATTTAA 23
          |||||
Db      498 AAAAAAAAACTTCATCATTTAA 520
```

```
RESULT 32
US-09-949-002-10470
; Sequence 10470, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10470
; LENGTH: 601.
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-10470
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 601;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTTCATCATTTAA 23
          |||||
Db      498 AAAAAAAAACTTCATCATTTAA 520
```

```
RESULT 33
US-09-270-767-7303
; Sequence 7303, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7303
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-7303
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 673;
```

```
Best Local Similarity 87.0%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 AAAAAAAAACTTCATCATTTAA 24
          |||||
Db      18 AAAAAATACCTTCACCAATTAA 40
```

```
RESULT 34
US-09-270-767-22585
; Sequence 22585, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22585
; LENGTH: 673
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-22585
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 673;
Best Local Similarity 87.0%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 AAAAAAAAACTTCATCATTTAA 24
          |||||
Db      18 AAAAAATACCTTCACCAATTAA 40
```

```
RESULT 35
US-09-601-198-154/c
; Sequence 154, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Casseil, Gail H.
; APPLICANT: Chen, Elison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Iefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 154
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-154
```

```
Query Match          72.8%; Score 18.2; DB 3; Length 1296;
Best Local Similarity 87.0%; Pred. No. 5.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTTCATCATTTAA 23
          |||||
Db      1031 AAAAAAACCTTCATTAATCAA 1009
```

```
RESULT 36
US-09-107-532A-62
; Sequence 62, Application US/09107532A
```

```

; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-5007
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3213 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1..3213
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-09-107-532A-62
Query Match 72.8%; Score 18.2; DB 3; Length 3213;
Best Local Similarity 87.0%; Pred. No. 5.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTTCATCATTTAA 24
Db 827 AAAAAAACCTTATCATTTCAA 849

RESULT 37
US-09-949-016-14432
; Sequence 14432, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

```

```

; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 14432
; LENGTH: 34172
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14432
Query Match 72.8%; Score 18.2; DB 3; Length 34172;
Best Local Similarity 87.0%; Pred. No. 6.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTTCATCATTTAA 24
Db 8190 AAAAAACATCTTCATTTAA 8212

RESULT 38
US-09-949-016-16150/C
; Sequence 16150, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16150
; LENGTH: 36620
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16150
Query Match 72.8%; Score 18.2; DB 3; Length 36620;
Best Local Similarity 87.0%; Pred. No. 6.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 23
Db 9309 AAAAAAACTCTCATTTAA 9287

RESULT 39
US-08-965-048-5/C
; Sequence 5, Application US/08965048
; Patent No. 6323244
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Fretner, Nelson
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
; TREATMENT OF NEUROPSYCHIATRIC DISORDERS
; FILE REFERENCE: 7853-093
; CURRENT APPLICATION NUMBER: US/08/965,048
; CURRENT FILING DATE: 1997-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 45716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-965-048-5

```

Query Match 72.8%; Score 18.2; DB 3; Length 45716;
Best Local Similarity 87.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 9160 AAAAAAAAAACCTTAATATATATA 9138

RESULT 40
US-08-965-048-6/c
; Sequence 6, Application US/08965048
; Patent No. 6323244
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Freimer, Nelson
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
; FILE REFERENCE: 7853-093
; CURRENT APPLICATION NUMBER: US/08/965,048
; CURRENT FILING DATE: 1997-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 45989
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-965-048-6

Query Match 72.8%; Score 18.2; DB 3; Length 45989;
Best Local Similarity 87.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 9160 AAAAAAAAAACCTTAATATATATA 9138

RESULT 41
US-09-949-016-12568
; Sequence 12568, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12568
; LENGTH: 56694
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12568

Query Match 72.8%; Score 18.2; DB 3; Length 56694;
Best Local Similarity 87.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 48172 AAAAAAAAAAGCGTTCATTTAA 48194

RESULT 42
US-09-949-016-15423
; Sequence 15423, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15423
; LENGTH: 56702
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15423

Query Match 72.8%; Score 18.2; DB 3; Length 56702;
Best Local Similarity 87.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 48172 AAAAAAAAAAGCGTTCATTTAA 48194

RESULT 43
US-09-949-016-15874
; Sequence 15874, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15874
; LENGTH: 60141
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)..(60141)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15874

Query Match 72.8%; Score 18.2; DB 3; Length 60141;
Best Local Similarity 87.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 37846 AAAAAAAAAAATTCATTTAA 37868

RESULT 44

```
US-09-949-002-693/c
; Sequence 693, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 693
; LENGTH: 63467
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-693

Query Match          72.8%; Score 18.2; DB 3; Length 63467;
Best Local Similarity 87.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTCATCATTTAA 23
Db 52942 AAAAAAAAAAAGCTTCATCATTTAA 52920

RESULT 45
US-09-949-016-17289/c
; Sequence 17289, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17289
; LENGTH: 64518
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17289

Query Match          72.8%; Score 18.2; DB 3; Length 64518;
Best Local Similarity 87.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTCATCATTTAA 23
Db 52942 AAAAAAAAAAAGCTTCATCATTTAA 52920

RESULT 46
US-09-949-002-843/c
; Sequence 843, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
```

```
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 843
; LENGTH: 64518
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-843

Query Match          72.8%; Score 18.2; DB 3; Length 64518;
Best Local Similarity 87.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTCATCATTTAA 23
Db 52942 AAAAAAAAAAAGCTTCATCATTTAA 52920

RESULT 47
US-09-949-016-16156/c
; Sequence 16156, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16156
; LENGTH: 77036
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...((77036)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16156

Query Match          72.8%; Score 18.2; DB 3; Length 77036;
Best Local Similarity 87.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTCATCATTTAA 23
Db 73262 AAAAAAAAAAAGCTTCATCATTTAA 73240

RESULT 48
US-09-949-016-15816/c
; Sequence 15816, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 15816
; LENGTH: 84839
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15816

```

```

Query Match          72.8%; Score 18.2; DB 3; Length 84839;
Best Local Similarity 87.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAACTTCATCATTTAA 23
Db      51252 AAAAAAAAACTTCATCATTTAA 51230

```

```

RESULT 49
US-09-949-016-11784
; Sequence 11784, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 11784
; LENGTH: 95318
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11784

```

```

Query Match          72.8%; Score 18.2; DB 3; Length 95318;
Best Local Similarity 87.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAACTTCATCATTTAA 23
Db      663 AAAAAAAAACTTCATCATTTAA 685

```

```

RESULT 50
US-09-949-016-13998
; Sequence 13998, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012

```

```

; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 13998
; LENGTH: 95318
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13998

```

```

Query Match          72.8%; Score 18.2; DB 3; Length 95318;
Best Local Similarity 87.0%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAACTTCATCATTTAA 23
Db      663 AAAAAAAAACTTCATCATTTAA 685

```

```

Search completed: December 14, 2005, 07:45:16
Job time : 67.5 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: December 14, 2005, 01:39:28 ; Search time 55.5 Seconds
(without alignments)

800.703 Million cell updates/sec

Title: US-10-681-773-10

Perfect score: 25

Sequence: 1 aaaaaaaaaaagcttgatcttc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/1 COMB.seq: *
2: /cgn2_6/prodata/1/ina/5 COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/H COMB.seq: *
6: /cgn2_6/prodata/1/ina/pCTUS COMB.seq: *
7: /cgn2_6/prodata/1/ina/PP COMB.seq: *
8: /cgn2_6/prodata/1/ina/RB COMB.seq: *
9: /cgn2_6/prodata/1/ina/backfills1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.2	80.8	627	3	US-09-270-767-9667 Sequence 9667, Ap
C 2	20.2	80.8	627	3	US-09-270-767-24949 Sequence 24949, A
C 3	19.8	79.2	93510	3	US-09-949-016-15095 Sequence 15095, A
C 4	19.2	76.8	187169	3	US-09-949-016-12776 Sequence 12776, A
C 5	19.2	76.8	191569	3	US-09-949-016-15940 Sequence 15940, A
C 6	18.8	75.2	945	3	US-09-243-681A-4161 Sequence 4161, Ap
C 7	18.8	75.2	1184	3	US-09-270-767-9472 Sequence 9472, Ap
C 8	18.8	75.2	1184	3	US-09-270-767-24754 Sequence 24754, A
C 9	18.8	75.2	1227	3	US-09-248-796A-6555 Sequence 6555, Ap
C 10	18.8	75.2	17527	3	US-09-949-016-13145 Sequence 13145, A
C 11	18.6	74.4	172	3	US-09-513-999C-20783 Sequence 20783, A
C 12	18.6	74.4	232	3	US-09-621-976-10815 Sequence 10815, A
C 13	18.6	74.4	601	3	US-09-949-016-11503 Sequence 112503, A
C 14	18.6	74.4	601	3	US-09-949-016-11503 Sequence 158191, A
C 15	18.6	74.4	661	3	US-09-328-111-69 Sequence 69, Appl
C 16	18.6	74.4	3073	2	US-07-688-352C-31 Sequence 31, Appl
C 17	18.6	74.4	3073	2	US-08-474-379C-31 Sequence 31, Appl
C 18	18.6	74.4	3073	3	US-09-146-249A-31 Sequence 31, Appl
C 19	18.6	74.4	3073	3	US-08-206-188B-31 Sequence 31, Appl
C 20	18.6	74.4	3073	6	PCT-US91-02714-30 Sequence 30, Appl
C 21	18.6	74.4	17056	3	US-09-245-041-3 Sequence 3, Appl
C 22	18.6	74.4	17056	3	US-09-358-055B-3 Sequence 3, Appl
C 23	18.6	74.4	17056	3	US-09-893-238-3 Sequence 3, Appl
C 24	18.6	74.4	36577	3	US-09-949-016-13176 Sequence 13176, A

25	18.6	74.4	52636	3	US-09-949-016-14839 Sequence 14839, A
26	18.6	74.4	77036	3	US-09-949-016-16156 Sequence 16156, A
27	18.6	74.4	88557	3	US-09-949-016-17028 Sequence 17028, A
28	18.6	74.4	101558	3	US-09-949-016-12243 Sequence 12243, A
C 29	18.6	74.4	118382	3	US-09-949-016-15996 Sequence 15996, A
C 30	18.6	74.4	118382	3	US-09-949-016-15997 Sequence 15997, A
C 31	18.6	74.4	278866	3	US-09-949-016-13922 Sequence 13922, A
C 32	18.6	74.4	278866	3	US-09-949-016-13923 Sequence 13923, A
C 33	18.6	74.4	278866	3	US-09-949-016-13924 Sequence 13924, A
C 34	18.6	74.4	278866	3	US-09-949-016-13925 Sequence 13925, A
C 35	18.6	74.4	278866	3	US-09-949-016-13926 Sequence 13926, A
C 36	18.6	74.4	278866	3	US-09-949-016-14699 Sequence 14699, A
C 37	18.6	74.4	278866	3	US-09-949-016-14700 Sequence 14700, A
C 38	18.6	74.4	278866	3	US-09-949-016-14701 Sequence 14701, A
C 39	18.6	74.4	278866	3	US-09-949-016-14702 Sequence 14702, A
C 40	18.6	74.4	278866	3	US-09-949-016-14703 Sequence 14703, A
C 41	18.6	74.4	390890	3	US-09-949-016-14720 Sequence 14720, A
C 42	18.4	73.6	601	3	US-09-949-016-55659 Sequence 55659, A
C 43	18.4	73.6	636591	3	US-09-949-016-11808 Sequence 11808, A
C 44	18.4	73.6	636591	3	US-09-949-016-13388 Sequence 13388, A
C 45	18.2	72.8	50	3	US-08-851-843A-102 Sequence 102, App
C 46	18.2	72.8	50	3	US-08-974-549A-587 Sequence 587, App
C 47	18.2	72.8	50	3	US-08-854-050-102 Sequence 102, App
C 48	18.2	72.8	50	3	US-09-430-323-102 Sequence 102, App
C 49	18.2	72.8	50	3	US-09-402-181B-587 Sequence 587, App
C 50	18.2	72.8	50	3	US-09-721-456-587 Sequence 587, App
C 51	18.2	72.8	50	3	US-09-766-253-102 Sequence 102, App
C 52	18.2	72.8	50	3	US-10-054-295-102 Sequence 102, App
C 53	18.2	72.8	50	3	US-09-438-486A-102 Sequence 102, App
C 54	18.2	72.8	52	3	US-09-057-328-1 Sequence 11, Appl
C 55	18.2	72.8	52	3	US-09-258-373-11 Sequence 11, Appl
C 56	18.2	72.8	52	3	US-09-091-590A-15 Sequence 15, Appl
C 57	18.2	72.8	52	3	US-09-228-455-1 Sequence 1, Appl
C 58	18.2	72.8	52	3	US-09-228-455-2 Sequence 2, Appl
C 59	18.2	72.8	52	3	US-09-276-533A-17 Sequence 17, Appl
C 60	18.2	72.8	52	3	US-10-158-735-17 Sequence 17, Appl
C 61	18.2	72.8	52	3	US-09-443-282B-12 Sequence 12, Appl
C 62	18.2	72.8	52	3	US-10-154-511-1 Sequence 1, Appl
C 63	18.2	72.8	52	3	US-10-154-511-2 Sequence 2, Appl
C 64	18.2	72.8	304	3	US-09-621-976-17988 Sequence 17988, A
C 65	18.2	72.8	601	3	US-09-949-016-13868 Sequence 52868, A
C 66	18.2	72.8	601	3	US-09-949-016-52788 Sequence 52788, A
C 67	18.2	72.8	601	3	US-09-949-016-96773 Sequence 96773, A
C 68	18.2	72.8	601	3	US-09-949-016-96779 Sequence 96779, A
C 69	18.2	72.8	601	3	US-09-949-016-96779 Sequence 96779, A
C 70	18.2	72.8	601	3	US-09-949-016-97045 Sequence 97045, A
C 71	18.2	72.8	601	3	US-09-949-016-97311 Sequence 97311, A
C 72	18.2	72.8	601	3	US-09-949-016-97577 Sequence 97577, A
C 73	18.2	72.8	601	3	US-09-949-016-97843 Sequence 97843, A
C 74	18.2	72.8	601	3	US-09-949-016-98109 Sequence 98109, A
C 75	18.2	72.8	601	3	US-09-949-016-98375 Sequence 98375, A
C 76	18.2	72.8	601	3	US-09-949-016-98641 Sequence 98641, A
C 77	18.2	72.8	601	3	US-09-949-016-98907 Sequence 98907, A
C 78	18.2	72.8	601	3	US-09-949-016-99173 Sequence 99173, A
C 79	18.2	72.8	601	3	US-09-949-016-99439 Sequence 99439, A
C 80	18.2	72.8	601	3	US-09-949-016-99705 Sequence 99705, A
C 81	18.2	72.8	601	3	US-09-949-016-99971 Sequence 99971, A
C 82	18.2	72.8	601	3	US-09-949-016-100237 Sequence 100237, A
C 83	18.2	72.8	601	3	US-09-949-016-100237 Sequence 100237, A
C 84	18.2	72.8	601	3	US-09-949-016-100769 Sequence 100769, A
C 85	18.2	72.8	601	3	US-09-949-016-101073 Sequence 101073, A
C 86	18.2	72.8	601	3	US-09-949-016-101339 Sequence 101339, A
C 87	18.2	72.8	601	3	US-09-949-016-101605 Sequence 101605, A
C 88	18.2	72.8	601	3	US-09-949-016-101871 Sequence 101871, A
C 89	18.2	72.8	601	3	US-09-949-016-102291 Sequence 102291, A
C 90	18.2	72.8	601	3	US-09-949-016-102431 Sequence 102431, A
C 91	18.2	72.8	669	3	US-09-328-111-342 Sequence 342, App
C 92	18.2	72.8	3479	3	US-08-714-918-7 Sequence 7, Appl
C 93	18.2	72.8	3479	3	US-09-265-315-7 Sequence 7, Appl
C 94	18.2	72.8	3479	3	US-09-265-315-7 Sequence 7, Appl
C 95	18.2	72.8	3479	3	US-09-266-417-7 Sequence 7, Appl
C 96	18.2	72.8	3479	3	US-09-528-709-7 Sequence 7, Appl
C 97	18.2	72.8	3479	3	US-09-527-745-7 Sequence 7, Appl

```
C 98 18.2 72.8 5924 3 US-08-956-171E-130
C 99 18.2 72.8 5924 3 US-08-781-986A-130
100 18.2 72.8 9139 3 US-09-322-478-22
101 18.2 72.8 9139 3 US-09-586-106D-22
102 18.2 72.8 9139 3 US-10-799-870-22
C 103 18.2 72.8 25992 3 US-09-949-016-17308
C 104 18.2 72.8 34502 3 US-09-949-016-14820
C 105 18.2 72.8 46319 3 US-09-949-016-15256
C 106 18.2 72.8 46323 3 US-09-949-016-13267
C 107 18.2 72.8 55298 3 US-09-491-356C-1
C 108 18.2 72.8 88490 3 US-09-949-016-12758
C 109 18.2 72.8 88736 3 US-09-949-016-14222
110 18.2 72.8 92581 3 US-09-949-016-12182
111 18.2 72.8 92581 3 US-09-949-016-16542
112 18.2 72.8 98439 3 US-09-949-016-13597
113 18.2 72.8 116652 3 US-09-949-016-13413
114 18.2 72.8 157866 3 US-09-949-016-12982
115 18.2 72.8 157866 3 US-09-949-016-12983
116 18.2 72.8 157866 3 US-09-949-016-12984
C 117 18.2 72.8 387902 3 US-09-949-016-14543
118 18.2 72.8 393753 3 US-09-949-016-14573
119 18.2 72.8 393753 3 US-09-949-016-14574
C 120 18.2 72.8 421883 3 US-09-949-016-12557
121 18.2 72.8 450395 3 US-09-949-016-15473
122 18.2 72.8 818128 3 US-09-949-016-14546
123 18.2 72.8 818128 3 US-09-949-016-14547
124 18.2 72.8 818128 3 US-09-949-016-14548
125 18.2 72.8 818128 3 US-09-949-016-14549
126 18.2 72.8 818128 3 US-09-949-016-14550
127 18.2 72.8 818128 3 US-09-949-016-14551
128 18.2 72.8 818128 3 US-09-949-016-14552
129 18.2 72.8 818128 3 US-09-949-016-14553
130 18.2 72.8 818128 3 US-09-949-016-14554
131 18.2 72.8 818128 3 US-09-949-016-14555
132 18.2 72.8 818128 3 US-09-949-016-14556
133 18.2 72.8 818128 3 US-09-949-016-14557
134 18.2 72.8 818128 3 US-09-949-016-14558
135 18.2 72.8 818128 3 US-09-949-016-14559
136 18.2 72.8 818128 3 US-09-949-016-14560
137 18.2 72.8 818128 3 US-09-949-016-14561
138 18.2 72.8 818128 3 US-09-949-016-14562
139 18.2 72.8 818128 3 US-09-949-016-14564
140 18.2 72.8 818128 3 US-09-949-016-14565
141 18.2 72.8 818128 3 US-09-949-016-14566
142 18.2 72.8 818128 3 US-09-949-016-14567
143 17.8 71.2 601 3 US-09-949-016-153451
144 17.8 71.2 601 3 US-09-949-016-153452
C 145 17.8 71.2 1336 3 US-09-949-016-204628
C 146 17.8 71.2 1336 3 US-09-257-179-13
147 17.8 71.2 2339 3 US-09-270-767-9189
148 17.8 71.2 2339 3 US-09-270-767-24471
C 149 17.8 71.2 43133 3 US-09-949-016-14911
150 17.8 71.2 122626 3 US-09-949-016-17524
```

ALIGNMENTS

```
Sequence 130, App
Sequence 130, App
Sequence 22, App1
Sequence 22, App1
Sequence 22, App1
Sequence 12, App1
Sequence 12, App1
Sequence 14820, A
Sequence 12526, A
Sequence 13267, A
Sequence 1, App1
Sequence 12758, A
Sequence 14282, A
Sequence 12182, A
Sequence 16542, A
Sequence 13597, A
Sequence 13413, A
Sequence 12982, A
Sequence 12983, A
Sequence 12984, A
Sequence 14543, A
Sequence 14573, A
Sequence 14574, A
Sequence 12557, A
Sequence 12557, A
Sequence 15473, A
Sequence 14546, A
Sequence 14547, A
Sequence 14548, A
Sequence 14549, A
Sequence 14550, A
Sequence 14551, A
Sequence 14552, A
Sequence 14553, A
Sequence 14554, A
Sequence 14555, A
Sequence 14556, A
Sequence 14557, A
Sequence 14558, A
Sequence 14559, A
Sequence 14560, A
Sequence 14561, A
Sequence 14562, A
Sequence 14564, A
Sequence 14565, A
Sequence 14566, A
Sequence 14567, A
Sequence 153451, A
Sequence 153452, A
Sequence 204628, A
Sequence 13, App1
Sequence 9189, App
Sequence 24471, A
Sequence 14911, A
Sequence 17524, A
```

```
RESULT 1
US-09-270-767-9667/c
; Sequence 9667, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 9667
; LENGTH: 627
; TYPE: DNA
```

```
; ORGANISM: Drosophila melanogaster
US-09-270-767-9667
Query Match 80.8%; Score 20.2; DB 3; Length 627;
Best Local Similarity 88.0%; Pred. No. 81;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Cy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 121 AACCAAACTAAGCTTGATCTTC 97
```

```
RESULT 2
US-09-270-767-24949/c
; Sequence 24949, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 24949
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24949
```

```
Query Match 80.8%; Score 20.2; DB 3; Length 627;
Best Local Similarity 88.0%; Pred. No. 81;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Cy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 121 AACCAAACTAAGCTTGATCTTC 97
```

```
RESULT 3
US-09-949-016-15095/c
; Sequence 15095, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; NUMBER OF SEQ ID NOS: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ For Windows Version 4.0
; SEQ ID NO 15095
; LENGTH: 93510
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(93510)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15095
```

```
Query Match 79.2%; Score 19.8; DB 3; Length 93510;
Best Local Similarity 91.3%; Pred. No. 1,4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Cy 1 AAAAAAAAACTAAGCTTGATCT 23
```


Db 18618 AAAAAAAAAAGTAAAGCTGATCT 18596

RESULT 4

US-09-949-016-12776/C
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12776
; LENGTH: 187169
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(187169)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

Query Match 76.8%; Score 19.2; DB 3; Length 187169;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 42190 AAAAAAAAACTAAAGCTTGATCT 42167

RESULT 5

US-09-949-016-15940/C
; Sequence 15940, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15940
; LENGTH: 191569
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(191569)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15940

Query Match 76.8%; Score 19.2; DB 3; Length 191569;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 42190 AAAAAAAAACTAAAGCTTGATCT 42167

RESULT 6

US-09-543-681A-4161/C
; Sequence 4161, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4161
; LENGTH: 945
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-4161

Query Match 75.2%; Score 18.8; DB 3; Length 945;
Best Local Similarity 90.9%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATC 22
Db 652 AACAAAACTAAAGCTTGATC 631

RESULT 7

US-09-270-767-9472/C
; Sequence 9472, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9472
; LENGTH: 1184
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9472

Query Match 75.2%; Score 18.8; DB 3; Length 1184;
Best Local Similarity 90.9%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATC 22
Db 60 AAAAAAAAACTAAAGCTTGATC 39

RESULT 8

US-09-270-767-24754/C
; Sequence 24754, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17

NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 24754
LENGTH: 1184
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-24754

Query Match
Best Local Similarity 90.9%; Score 18.8; DB 3; Length 1184;
Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATC 22
|||||
Db 60 AAAAAAAAACTGAAGCTTATC 39
|||||

RESULT 9
US-09-248-796A-6555/c
Sequence 6555, Application US/09248796A
Patent No. 6747137

GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO: 6555
LENGTH: 1227
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-6555

Query Match
Best Local Similarity 90.9%; Score 18.8; DB 3; Length 1227;
Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAAAAAAACTAAGCTTGATCTT 24
|||||
Db 632 AAAAAAAAACTAAGTTTGATCTT 611
|||||

RESULT 10
US-09-949-016-13145/c
Sequence 13145, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13145
LENGTH: 17527
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature

LOCATION: (1)...(17527)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13145

Query Match
Best Local Similarity 90.9%; Score 18.8; DB 3; Length 17527;
Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
Db 15906 AAAAAAAAACTGAAGCTGATGTC 15885
|||||

RESULT 11
US-09-513-999C-20783/c
Sequence 20783, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO: 20783
LENGTH: 172
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 149
OTHER INFORMATION: w=a or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 161
OTHER INFORMATION: w=a or t
US-09-513-999C-20783

Query Match
Best Local Similarity 84.0%; Score 18.6; DB 3; Length 172;
Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
Db 78 AAAAAAAAACTGAAGCTGATGATC 54
|||||

RESULT 12
US-09-621-976-10815/c
Sequence 10815, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSER.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO: 10815
LENGTH: 232
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-10815

Query Match 74.4%; Score 18.6; DB 3; Length 232;
Best Local Similarity 84.0%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTTC 25
|||||
Db 230 AAAAAAAAAAAAGCTTGTCCTTC 206

RESULT 13
US-09-949-016-112503
; Sequence 112503, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 112503
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-112503

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTTC 25
|||||
Db 466 AAAAAAAAAAAAGTATGAGCTTC 490

RESULT 14
US-09-949-016-158191
; Sequence 158191, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 158191
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-158191

Query Match 74.4%; Score 18.6; DB 3; Length 601;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTTC 25

Db 159 AAAAAAAAAATTAAGCTTGTCCTTC 183
|||||

RESULT 15
US-09-328-111-69/C
; Sequence 69, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Axtle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Carino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 69
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(661)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-69

Query Match 74.4%; Score 18.6; DB 3; Length 661;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTTC 25
|||||
Db 438 AAAAAAAAAAAAGCTTGATCTTC 414

RESULT 16
US-07-688-352C-31
; Sequence 31, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/688,352C
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
US-07-688-352C-31

Query Match 74.4%; Score 18.6; DB 2; Length 3073;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 1547 AAAAAAAAAAAGCTTGATTC 1571

RESULT 17
US-08-474-379C-31
Sequence 31, Application US/08474379C
Patent No. 5977305
GENERAL INFORMATION:
APPLICANT: Migler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
TITLE OF INVENTION: PROCESSES
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,379C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,188
FILING DATE: 01-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/688,352
FILING DATE: 19-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.

REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 27866/32771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1109
US-08-474-379C-31

Query Match 74.4%; Score 18.6; DB 2; Length 3073;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 1547 AAAAAAAAAAAGCTTGATTC 1571

RESULT 18
US-09-146-249A-31
Sequence 31, Application US/09146249A
Patent No. 6069240
GENERAL INFORMATION:
APPLICANT: Migler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/146,249A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
US-09-146-249A-31

Query Match 74.4%; Score 18.6; DB 3; Length 3073;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTTC 25
Db 1547 AAAAAAAAAAAAGCTGTATTC 1571

RESULT 19

US-08-206-188B-31
; Sequence 31, Application US/08206188B
; Patent No. 6100025
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 84
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,188B
; FILING DATE: 01-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3073 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1111
; US-08-206-188B-31

Query Match 74.4%; Score 18.6; DB 3; Length 3073;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTTC 25
Db 1547 AAAAAAAAAAAAGCTGTATTC 1571

RESULT 20

PCT-US91-02714-30
; Sequence 30, Application PC/TUS9102714
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.

; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/02714
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3073 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1111
; PCT-US91-02714-30

Query Match 74.4%; Score 18.6; DB 6; Length 3073;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTTC 25
Db 1547 AAAAAAAAAAAAGCTGTATTC 1571

RESULT 21

US-09-245-041-3/c
; Sequence 3, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Wigler, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
; FILE REFERENCE: 7853-136
; CURRENT APPLICATION NUMBER: US/09/245,041
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/093,630
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/104,978
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 17056

TYPE: DNA
ORGANISM: Mus musculus
US-09-245-041-3

Query Match 74.4%; Score 18.6; DB 3; Length 17056;
Best Local Similarity 84.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
DB 394 AAAAAAAAAATTAAGCAAGATCTTC 370

RESULT 22
US-09-358-055B-3/c
Sequence 3, Application US/09358055B
Patent No. 6713277
GENERAL INFORMATION:
APPLICANT: Nagle, D.L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF BODY WEIGHT DISORDERS INCLUDING
FILE REFERENCE: 7853-151
CURRENT APPLICATION NUMBER: US/09/358,055B
PRIOR FILING DATE: 1999-07-21
PRIOR APPLICATION NUMBER: 09/245,041
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 153
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 17056
TYPE: DNA
ORGANISM: Mus musculus
US-09-358-055B-3

Query Match 74.4%; Score 18.6; DB 3; Length 17056;
Best Local Similarity 84.0%; Pred. No. 3.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
DB 394 AAAAAAAAAATTAAGCAAGATCTTC 370

RESULT 23
US-09-893-238-3/c
Sequence 3, Application US/09893238
Patent No. 6727348
GENERAL INFORMATION:
APPLICANT: Moore, K.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND
TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
FILE REFERENCE: 7853-237
CURRENT APPLICATION NUMBER: US/09/893,238
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: 09/245,041
PRIOR FILING DATE: 1999-02-05
PRIOR APPLICATION NUMBER: 60/093,630
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: 60/104,978
PRIOR FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 129
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 17056
TYPE: DNA
ORGANISM: Mus musculus
US-09-893-238-3

Query Match 74.4%; Score 18.6; DB 3; Length 17056;
Best Local Similarity 84.0%; Pred. No. 3.7e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
DB 394 AAAAAAAAAATTAAGCAAGATCTTC 370

RESULT 24
US-09-949-016-13176
Sequence 13176, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13176
LENGTH: 36577
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(36577)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13176

Query Match 74.4%; Score 18.6; DB 3; Length 36577;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
DB 20708 AAAAAAAAAAAAGCTTGATCTTC 20732

RESULT 25
US-09-949-016-14839
Sequence 14839, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14839
LENGTH: 52636
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(52636)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14839

```
Query Match          74.4%; Score 18.6; DB 3; Length 52636;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 48937 AAAAAAAAAAAGATGAGCTTC 48961

RESULT 26
US-09-949-016-16156
; Sequence 16156, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16156
; LENGTH: 77036
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...((77036))
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16156

Query Match          74.4%; Score 18.6; DB 3; Length 77036;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 28376 AAAAAAAAAATTAGCTTGCTGCTTC 28400

RESULT 27
US-09-949-016-17028
; Sequence 17028, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17028
; LENGTH: 88557
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)...(88557)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17028

Query Match          74.4%; Score 18.6; DB 3; Length 88557;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 79853 AAAAAAAAAAAGTTGATATTC 79877

RESULT 28
US-09-949-016-12243
; Sequence 12243, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12243
; LENGTH: 101558
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(101558)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12243

Query Match          74.4%; Score 18.6; DB 3; Length 101558;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 52912 AAAAAAAAAATTAGCTTGCTGCTTC 52936

RESULT 29
US-09-949-016-15996/C
; Sequence 15996, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15996
; LENGTH: 118382
; TYPE: DNA
```

```
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(118382)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15996
```

```
Query Match
Best Local Similarity 74.4%; Score 18.6; DB 3; Length 118382;
                          84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 50382 AAAAAAAAAAACTGTTCTTC 50358
```

```
RESULT 30
US-09-949-016-15997/C
Sequence 15997, Application US/09949016
Patent No. 6812339
```

```
GENERAL INFORMATION:
```

```
APPLICANT: VENTER, J. Craig et al.
```

```
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
FILE REFERENCE: CLO01307 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
```

```
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```

```
PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 15997
```

```
LENGTH: 118382
```

```
TYPE: DNA
```

```
ORGANISM: Human
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (1)...(118382)
```

```
OTHER INFORMATION: n = A,T,C or G
```

```
US-09-949-016-15997
```

```
Query Match
Best Local Similarity 74.4%; Score 18.6; DB 3; Length 118382;
                          84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 50382 AAAAAAAAAAACTGTTCTTC 50358
```

```
RESULT 31
US-09-949-016-13922
Sequence 13922, Application US/09949016
Patent No. 6812339
```

```
GENERAL INFORMATION:
```

```
APPLICANT: VENTER, J. Craig et al.
```

```
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
FILE REFERENCE: CLO01307 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
```

```
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```

```
PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 13922
LENGTH: 278866
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(278866)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13922
```

```
Query Match
Best Local Similarity 74.4%; Score 18.6; DB 3; Length 278866;
                          84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTGTTCTTC 203636
```

```
RESULT 32
US-09-949-016-13923
Sequence 13923, Application US/09949016
Patent No. 6812339
```

```
GENERAL INFORMATION:
```

```
APPLICANT: VENTER, J. Craig et al.
```

```
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
FILE REFERENCE: CLO01307 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
```

```
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```

```
PRIOR FILING DATE: 2000-09-08
```

```
NUMBER OF SEQ ID NOS: 207012
```

```
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 13923
```

```
LENGTH: 278866
```

```
TYPE: DNA
```

```
ORGANISM: Human
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (1)...(278866)
```

```
OTHER INFORMATION: n = A,T,C or G
```

```
US-09-949-016-13923
```

```
Query Match
Best Local Similarity 74.4%; Score 18.6; DB 3; Length 278866;
                          84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTGTTCTTC 203636
```

```
RESULT 33
US-09-949-016-13924
Sequence 13924, Application US/09949016
Patent No. 6812339
```

```
GENERAL INFORMATION:
```

```
APPLICANT: VENTER, J. Craig et al.
```

```
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
FILE REFERENCE: CLO01307 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
CURRENT APPLICATION NUMBER: US/09/949,016
```

```
PRIOR FILING DATE: 2000-04-14
```

```
PRIOR APPLICATION NUMBER: 60/241,755
```

```
PRIOR FILING DATE: 2000-10-20
```

```
PRIOR APPLICATION NUMBER: 60/237,768
```

```
PRIOR FILING DATE: 2000-10-03
```

```
PRIOR APPLICATION NUMBER: 60/231,498
```



```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13924
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13924

Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636

RESULT 34
US-09-949-016-13925
; Sequence 13925, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13925
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13925

Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636

RESULT 35
US-09-949-016-13926
; Sequence 13926, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13926
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13926

Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636

RESULT 36
US-09-949-016-14699
; Sequence 14699, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14699
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14699

Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636

RESULT 37
US-09-949-016-14700
; Sequence 14700, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14700
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14700
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636
```

```
RESULT 38
US-09-949-016-14701
; Sequence 14701, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14701
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14701
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636
```

```
RESULT 39
US-09-949-016-14702
; Sequence 14702, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14702
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14702
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636
```

```
RESULT 40
US-09-949-016-14703
; Sequence 14703, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 14703
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14703
```

```
Query Match          74.4%; Score 18.6; DB 3; Length 278866;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 203612 AAAAAAAAAAACTTGTTCTTC 203636
```

```
RESULT 41
US-09-949-016-14720/C
; Sequence 14720, Application US/09949016
; Patent No. 6812339
```

```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14720
; LENGTH: 390890
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(390890)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14720

Query Match          74.4%; Score 18.6; DB 3; Length 390890;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
DB 103001 AAAAAAAAACTAAGCTTGATCTTC 102977

RESULT 42
US-09-949-016-56569
; Sequence 56569, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56569
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-56569

Query Match          73.6%; Score 18.4; DB 3; Length 601;
Best Local Similarity 95.0%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGAT 21
DB 258 AAAAAAAAACTAAGCTTGAT 277

RESULT 43
US-09-949-016-11808
; Sequence 11808, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11808
; LENGTH: 636591
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(636591)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11808

Query Match          73.6%; Score 18.4; DB 3; Length 636591;
Best Local Similarity 95.0%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGAT 21
DB 71338 AAAAAAAAACTAAGCTTGAT 71357

RESULT 44
US-09-949-016-13388
; Sequence 13388, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13388
; LENGTH: 636591
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(636591)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13388

Query Match          73.6%; Score 18.4; DB 3; Length 636591;
Best Local Similarity 95.0%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGAT 21
DB 71338 AAAAAAAAACTAAGCTTGAT 71357

RESULT 45
US-08-851-843A-102/c
```

```
; Sequence 102, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cecch, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin H.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809e1 Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,843A
; FILING DATE: 06-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-851-843A-102

Query Match 72.8%; Score 18.2; DB 3; Length 50;
Best Local Similarity 87.0%; Pred No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTAAGCTTGATCT 23
    |||||
Db 49 AAAAAAAAAAGCTTGAGCT 27
```

```
RESULT 46
US-08-974-549A-587/c
; Sequence 587, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cecch, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
```

```
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin H.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 587:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..50
; OTHER INFORMATION: /note= "Q-T primer"
; US-08-974-549A-587

Query Match 72.8%; Score 18.2; DB 3; Length 50;
```

Best Local Similarity 87.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 23
49 AAAAAAAAAAAGCTTGAGCT 27
Db

RESULT 47

US-08-854-050-102/c
; Sequence 102, Application US/08854050
; Patent No. 6261836

GENERAL INFORMATION:

APPLICANT: Cecch, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin H.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 102:

SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-854-050-102

Query Match 72.8%; Score 18.2; DB 3; Length 50;
Best Local Similarity 87.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 23

49 AAAAAAAAAAAGCTTGAGCT 27

Db

RESULT 48

US-09-430-323-102/c
; Sequence 102, Application US/09430323
; Patent No. 6309867

GENERAL INFORMATION:

APPLICANT: Cecch, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin H.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6309867el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,323
FILING DATE: 29-Oct-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997

APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997

APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 102:

SEQUENCE CHARACTERISTICS:

LENGTH: 50 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

SEQUENCE DESCRIPTION: SEQ ID NO: 102:

US-09-430-323-102

Query Match 72.8%; Score 18.2; DB 3; Length 50;
Best Local Similarity 87.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 23
49 AAAAAAAAAAAGCTTGAGCT 27
Db

RESULT 49
US-09-402-181B-587/C
Sequence 587, Application US/09402181B
Patent No. 6610839
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 633
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/402,181B
FILING DATE: 29-Sep-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ausenbus, Scott L.
REGISTRATION NUMBER: 42,271
REFERENCE/DOCKET NUMBER: 015389-002620US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 587:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..50
OTHER INFORMATION: /note="Q-T primer"
SEQUENCE DESCRIPTION: SEQ ID NO: 587:
US-09-402-181B-587
Query Match 72.8%; Score 18.2; DB 3; Length 50;
Best Local Similarity 87.0%; Pred. No. 4,3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTGATCT 23
Db 49 AAAAAAAAAAAGCTTGAGCT 27
RESULT 50
US-09-721-456-587/C
Sequence 587, Application US/09721456
Patent No. 6617110
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
Lingner, Joachim
Nakamura, Toru
Chapman, Karen B.
Morin, Gregg B.
Harley, Calvin B.
Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/721,456
FILING DATE: 22-No. 6617110-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 587:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

```

FEATURE:
;
; NAME/KEY:
; LOCATION: 1..50
; OTHER INFORMATION: /note="O-T primer"
; SEQUENCE DESCRIPTION: SEQ ID NO: 587:
US-09-721-456-587

```

```

Query Match          72.8%; Score 18.2; DB 3; Length 50;
Best Local Similarity 87.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1  AAAAAAAAACTAAAGCTGATCT 23
Db      49  AAAAAAAAAAAAGCTGAGCT 27

```

Search completed: December 14, 2005, 07:45:30
 Job time : 69.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 ; Search time 373.2 Seconds

(without alignment)
553.951 Million cell updates/sec

Title: US-10-681-773-1

Perfect score: 25
Sequence: 1 aaaaaaaaaatcgcaacaatc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

- Published Applications_NA_Main:*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US03B_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	7	US-10-681-773-1
2	25	100.0	25473	8	US-10-719-993-7034
3	24.6	98.4	33	9	US-10-891-260-1793
4	24.6	98.4	552	5	US-10-027-632-78265
5	24.6	98.4	552	5	US-10-027-632-78266
6	24.6	98.4	552	5	US-10-027-632-300721
7	24.6	98.4	552	6	US-10-027-632-300722
8	24.6	98.4	552	6	US-10-027-632-78265
9	24.6	98.4	552	6	US-10-027-632-78266
10	24.6	98.4	552	6	US-10-027-632-300721
11	24.6	98.4	552	6	US-10-027-632-300722
12	24.6	93.6	25	7	US-10-681-773-2
13	23.4	93.6	621	4	US-09-925-065A-887427
14	23.4	93.6	621	4	US-09-925-065A-887428
15	23.4	93.6	621	4	US-09-925-065A-887429
16	23.4	93.6	621	4	US-09-925-065A-887430
17	23.4	93.6	689	4	US-09-925-065A-831477
18	23.4	93.6	142318	5	US-10-087-192-1504
19	23	92.0	25	7	US-10-681-773-1
20	22	88.0	25	7	US-10-681-773-26
21	22	88.0	58038	8	US-10-741-600-17942
22	22	88.0	333811	8	US-10-741-600-17681
23	21.8	87.2	458	5	US-10-027-632-57229

24	21.8	87.2	458	5	US-10-027-632-57230	Sequence 57230, A
25	21.8	87.2	458	5	US-10-027-632-309952	Sequence 309952, A
26	21.8	87.2	458	5	US-10-027-632-309953	Sequence 309953, A
27	21.8	87.2	458	6	US-10-027-632-57229	Sequence 57229, A
28	21.8	87.2	458	6	US-10-027-632-57230	Sequence 57230, A
29	21.8	87.2	458	6	US-10-027-632-309952	Sequence 309952, A
30	21.8	87.2	458	6	US-10-027-632-309953	Sequence 309953, A
31	21.8	87.2	539	4	US-09-925-065A-384901	Sequence 384901, A
32	21.8	87.2	539	5	US-10-027-632-4931	Sequence 4931, Ap
33	21.8	87.2	539	6	US-10-027-632-4931	Sequence 4931, Ap
34	21.8	87.2	540	5	US-10-027-632-189916	Sequence 189916, A
35	21.8	87.2	540	5	US-10-027-632-189917	Sequence 189917, A
36	21.8	87.2	540	6	US-10-027-632-189916	Sequence 189916, A
37	21.8	87.2	540	6	US-10-027-632-189917	Sequence 189917, A
38	21.8	87.2	551	4	US-09-925-065A-184309	Sequence 184309, A
39	21.8	87.2	590	4	US-09-925-065A-702224	Sequence 702224, A
40	21.8	87.2	590	4	US-09-925-065A-702225	Sequence 702225, A
41	21.8	87.2	590	4	US-09-925-065A-702226	Sequence 702226, A
42	21.8	87.2	636	4	US-09-925-065A-726714	Sequence 726714, A
43	21.8	87.2	636	4	US-09-925-065A-726715	Sequence 726715, A
44	21.8	87.2	636	4	US-09-925-065A-726716	Sequence 726716, A
45	21.8	87.2	640	5	US-10-027-632-70791	Sequence 70791, A
46	21.8	87.2	640	5	US-10-027-632-319807	Sequence 319807, A
47	21.8	87.2	640	6	US-10-027-632-319807	Sequence 319807, A
48	21.8	87.2	640	6	US-10-027-632-150512	Sequence 150512, A
49	21.8	87.2	714	6	US-10-027-632-150512	Sequence 150512, A
50	21.8	87.2	714	6	US-09-925-065A-691495	Sequence 691495, A
51	21.8	87.2	840	4	US-09-925-065A-691496	Sequence 691496, A
52	21.8	87.2	840	4	US-09-925-065A-691497	Sequence 691497, A
53	21.8	87.2	840	5	US-10-027-632-38334	Sequence 38334, A
54	21.8	87.2	864	6	US-10-027-632-38334	Sequence 38334, A
55	21.8	87.2	902	6	US-10-027-632-4699	Sequence 4699, Ap
56	21.8	87.2	902	5	US-10-027-632-4699	Sequence 4699, Ap
57	21.8	87.2	902	5	US-10-027-632-4699	Sequence 4699, Ap
58	21.8	87.2	902	6	US-10-027-632-4699	Sequence 4699, Ap
59	21.8	87.2	902	6	US-10-027-632-4699	Sequence 4699, Ap
60	21.8	87.2	6162	7	US-10-240-589C-124	Sequence 124, App
61	21.8	87.2	6226	3	US-09-764-847-1961	Sequence 1961, App
62	21.8	87.2	6226	5	US-10-092-154-1961	Sequence 1961, App
63	21.8	87.2	7145	7	US-10-240-589C-43	Sequence 43, Appl
64	21.8	87.2	7928	6	US-10-311-455-68	Sequence 68, Appl
65	21.8	87.2	7928	7	US-10-322-281-308	Sequence 308, Appl
66	21.8	87.2	180227	7	US-10-322-281-308	Sequence 308, Appl
67	21.4	85.6	25	7	US-10-681-773-17	Sequence 17, Appl
68	21.4	85.6	539	4	US-09-925-065A-384902	Sequence 384902, A
69	21.4	85.6	1710	6	US-10-311-455-2358	Sequence 2358, Ap
70	21.4	85.6	3673778	6	US-10-312-841-2	Sequence 2, Appl
71	21	84.0	25	7	US-10-681-773-57	Sequence 57, Appl
72	21	84.0	230	3	US-09-854-867-324	Sequence 324, App
73	21	84.0	230	9	US-10-786-970A-324	Sequence 324, App
74	20.8	83.2	422	4	US-09-925-065A-351562	Sequence 351562, A
75	20.8	83.2	544	4	US-09-925-065A-110877	Sequence 110877, A
76	20.8	83.2	544	8	US-10-363-345A-12417	Sequence 12417, A
77	20.8	83.2	544	8	US-10-363-345A-12418	Sequence 12418, A
78	20.8	83.2	544	9	US-10-363-483A-12417	Sequence 12417, A
79	20.8	83.2	544	9	US-10-363-483A-12418	Sequence 12418, A
80	20.8	83.2	545	4	US-09-925-065A-93633	Sequence 93633, A
81	20.8	83.2	555	4	US-09-925-065A-108610	Sequence 108610, A
82	20.8	83.2	570	4	US-09-925-065A-254576	Sequence 254576, A
83	20.8	83.2	570	4	US-09-925-065A-254577	Sequence 254577, A
84	20.8	83.2	570	4	US-09-925-065A-254578	Sequence 254578, A
85	20.8	83.2	587	8	US-10-425-115-93990	Sequence 93990, A
86	20.8	83.2	594	4	US-09-925-065A-178544	Sequence 178544, A
87	20.8	83.2	635	4	US-09-925-065A-699791	Sequence 699791, A
88	20.8	83.2	635	4	US-09-925-065A-699792	Sequence 699792, A
89	20.8	83.2	635	4	US-09-925-065A-699793	Sequence 699793, A
90	20.8	83.2	635	4	US-09-925-065A-715390	Sequence 715390, A
91	20.8	83.2	722	8	US-10-363-345A-21751	Sequence 21751, A
92	20.8	83.2	722	8	US-10-363-345A-21752	Sequence 21752, A
93	20.8	83.2	722	9	US-10-363-483A-21751	Sequence 21751, A
94	20.8	83.2	722	9	US-10-363-483A-21752	Sequence 21752, A
95	20.8	83.2	730	8	US-10-363-345A-1811	Sequence 1811, Ap
96	20.8	83.2	730	8	US-10-363-345A-1812	Sequence 1812, Ap

```
C 97 20.8 83.2 730 9 US-10-363-483A-1811 Sequence 1811, Ap
98 20.8 83.2 730 5 US-10-363-483A-1812 Sequence 1812, Ap
99 20.8 83.2 1095 5 US-10-027-632-117597 Sequence 117597,
100 20.8 83.2 1095 5 US-10-027-632-117598 Sequence 117598,
101 20.8 83.2 1095 6 US-10-027-632-117597 Sequence 117597,
102 20.8 83.2 1095 6 US-10-027-632-117598 Sequence 117598,
103 20.8 83.2 1834 6 US-10-356-820-5 Sequence 5, Appl1
104 20.8 83.2 1834 6 US-10-260-238-1625 Sequence 1625, Ap
105 20.8 83.2 5763 6 US-10-311-455-155 Sequence 155, App
106 20.8 83.2 6127 6 US-10-311-455-158 Sequence 158, App
107 20.8 83.2 6221 6 US-10-221-714A-225 Sequence 225, App
108 20.8 83.2 8951 6 US-10-311-455-768 Sequence 768, App
109 20.8 83.2 13038 6 US-10-311-455-1248 Sequence 1248, Ap
110 20.8 83.2 14542 7 US-10-221-613-105 Sequence 105, App
111 20.8 83.2 15872 7 US-10-221-714A-241 Sequence 241, App
112 20.8 83.2 17968 8 US-10-602-494-151 Sequence 151, App
113 20.8 83.2 17968 8 US-10-602-494-293 Sequence 293, App
114 20.8 83.2 32132 3 US-09-764-877-2308 Sequence 2308, Ap
115 20.8 83.2 32132 3 US-09-764-877-2308 Sequence 2308, Ap
116 20.8 83.2 32132 6 US-10-242-515-2308 Sequence 2308, Ap
117 20.8 83.2 32169 5 US-09-764-847-1963 Sequence 1963, Ap
118 20.8 83.2 32169 5 US-10-092-154-1963 Sequence 1963, Ap
119 20.8 83.2 403035 7 US-10-092-154-1963 Sequence 1963, Ap
120 20.4 81.6 25 7 US-10-681-773-27 Sequence 27, Appl
121 20.4 81.6 88 3 US-09-783-590-6475 Sequence 6475, Ap
122 20.4 81.6 322 8 US-10-425-115-91122 Sequence 91122, A
123 20.4 81.6 376 8 US-10-357-930-61224 Sequence 61224, A
124 20.4 81.6 483 5 US-10-027-632-316127 Sequence 316127,
125 20.4 81.6 483 6 US-10-027-632-316127 Sequence 316127,
126 20.4 81.6 501 4 US-09-925-065A-154094 Sequence 154094,
127 20.4 81.6 518 4 US-09-925-065A-533029 Sequence 533029,
128 20.4 81.6 518 4 US-09-925-065A-533029 Sequence 533029,
129 20.4 81.6 554 4 US-09-925-065A-535614 Sequence 535614,
130 20.4 81.6 567 4 US-09-925-065A-287580 Sequence 287580,
131 20.4 81.6 581 4 US-09-925-065A-802052 Sequence 802052,
132 20.4 81.6 585 8 US-10-363-345A-6463 Sequence 6463, Ap
133 20.4 81.6 585 8 US-10-363-345A-6464 Sequence 6464, Ap
134 20.4 81.6 585 9 US-10-363-483A-6464 Sequence 6464, Ap
135 20.4 81.6 605 4 US-09-925-065A-390820 Sequence 390820,
136 20.4 81.6 605 4 US-09-925-065A-390821 Sequence 390821,
137 20.4 81.6 605 4 US-09-925-065A-390822 Sequence 390822,
138 20.4 81.6 605 4 US-09-925-065A-390823 Sequence 390823,
139 20.4 81.6 610 8 US-10-357-930-539823 Sequence 539823, A
140 20.4 81.6 639 4 US-09-925-065A-715589 Sequence 715589,
141 20.4 81.6 800 5 US-10-027-632-156259 Sequence 156259,
142 20.4 81.6 800 5 US-10-027-632-156259 Sequence 156259,
143 20.4 81.6 800 6 US-10-027-632-156259 Sequence 156259,
144 20.4 81.6 800 6 US-10-027-632-156259 Sequence 156259,
145 20.4 81.6 850 8 US-10-363-345A-15245 Sequence 35245, A
146 20.4 81.6 850 8 US-10-363-345A-15245 Sequence 35245, A
147 20.4 81.6 850 9 US-10-363-483A-35245 Sequence 35245, A
148 20.4 81.6 850 9 US-10-363-483A-35246 Sequence 35246, A
149 20.4 81.6 904 8 US-10-363-345A-25351 Sequence 25351, A
150 20.4 81.6 904 8 US-10-363-345A-25351 Sequence 25351, A
```

ALIGNMENTS

```
RESULT 1
US-10-681-773-1
; Sequence 1, Application US/10681773
; Publication No. US20040146690A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 1
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-1
```

```
Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 1 AAAAAAAAAATCGCAACCAATCT 25
```

```
RESULT 2
US-10-719-993-7034
; Sequence 7034, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
```

```
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
```

```
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7033
; LENGTH: 25473
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-719-993-7034
```

```
Query Match 100.0%; Score 25; DB 8; Length 25473;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 2923 AAAAAAAAAATCGCAACCAATCT 2947
```

```
RESULT 3
US-10-891-260-1793
; Sequence 1793, Application US/10891260
; Publication No. US20050227244A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Affymetrix, Inc.
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
```

```
; TITLE OF INVENTION: Methods of Analysis of Human Polymorphisms
; FILE REFERENCE: 3522.3
; CURRENT APPLICATION NUMBER: US/10/891,260
; CURRENT FILING DATE: 2004-07-13
```

```
; PRIOR APPLICATION NUMBER: 10/681,773
; PRIOR FILING DATE: 2003-10-07
; NUMBER OF SEQ ID NOS: 10244
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1793
; LENGTH: 33
; TYPE: DNA
```

```
; ORGANISM: homo sapien
US-10-891-260-1793
```

```
Query Match 98.4%; Score 24.6; DB 9; Length 33;
Best Local Similarity 96.0%; Pred. No. 51;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

PRIOR FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23

RESULT 7
US-10-027-632-300722

```
; Sequence 300722, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 325720
; SEQ ID NO 300722
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-300722
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 552;
Best Local Similarity 96.0%; Pred. No. 69;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 442 AAAAAAAAAATCGCARACCAATCT 466
```

```
RESULT 8
US-10-027-632-78265
; Sequence 78265, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 325720
; SEQ ID NO 78265
; LENGTH: 552
```

```
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-78265
```

```
Query Match          98.4%; Score 24.6; DB 6; Length 552;
Best Local Similarity 96.0%; Pred. No. 69;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 442 AAAAAAAAAATCGCARACCAATCT 466
```

```
RESULT 9
US-10-027-632-78266
; Sequence 78266, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 325720
; SEQ ID NO 78266
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-78266
```

```
Query Match          98.4%; Score 24.6; DB 6; Length 552;
Best Local Similarity 96.0%; Pred. No. 69;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 442 AAAAAAAAAATCGCARACCAATCT 466
```

```
RESULT 10
US-10-027-632-300721
; Sequence 300721, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
```

```

; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300721
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-300721

Query Match          98.4%; Score 24.6; DB 6; Length 552;
Best Local Similarity 96.0%; Pred. No. 69;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25
Db 442 AAAAAAAAAATCGCACAACTCT 466

RESULT 11
US-10-027-632-300722
; Sequence 300722, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300722
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-300722
```

```

Query Match          98.4%; Score 24.6; DB 6; Length 552;
Best Local Similarity 96.0%; Pred. No. 69;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25
Db 442 AAAAAAAAAATCGCACAACTCT 466

RESULT 12
US-10-681-773-2
; Sequence 2, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 2
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-2

Query Match          93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 12e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25
Db 1 AAAAAAAAAATCGCACAACTCT 25
```

```

RESULT 13
US-09-925-065A-887427
; Sequence 887427, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887427
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-887427

Query Match          93.6%; Score 23.4; DB 4; Length 621;
```

Best Local Similarity 96.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||
Db 443 AAAAAAAAAATCACCACAAATCT 467

RESULT 14
US-09-925-065A-887428

/ Sequence 887428, Application US/09925065A
/ Publication No. US20050228172A9

/ GENERAL INFORMATION:

/ APPLICANT: Wang, David G.

/ TITLE OF INVENTION: Identification and Mapping of Single

/ FILE REFERENCE: 108827.135

/ CURRENT APPLICATION NUMBER: US/09/925,065A

/ PRIOR FILING DATE: 2001-08-08

/ PRIOR APPLICATION NUMBER: US 60/243,096

/ PRIOR FILING DATE: 2000-10-24

/ PRIOR APPLICATION NUMBER: US 60/252,147

/ PRIOR FILING DATE: 2000-11-20

/ PRIOR APPLICATION NUMBER: US 60/250,092

/ PRIOR FILING DATE: 2000-11-30

/ PRIOR APPLICATION NUMBER: US 60/261,766

/ PRIOR FILING DATE: 2001-01-16

/ PRIOR APPLICATION NUMBER: US 60/289,846

/ PRIOR FILING DATE: 2001-05-09

/ NUMBER OF SEQ ID NOS: 957086

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 887428

/ LENGTH: 621

/ TYPE: DNA

/ ORGANISM: Homo sapiens

US-09-925-065A-887428

Query Match 93.6%; Score 23.4; DB 4; Length 621;
Best Local Similarity 96.0%; Pred. No. 1.7e+02;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||
Db 443 AAAAAAAAAATCACCACAAATCT 467

RESULT 15

US-09-925-065A-887429

/ Sequence 887429, Application US/09925065A
/ Publication No. US20050228172A9

/ GENERAL INFORMATION:

/ APPLICANT: Wang, David G.

/ TITLE OF INVENTION: Identification and Mapping of Single

/ FILE REFERENCE: 108827.135

/ CURRENT APPLICATION NUMBER: US/09/925,065A

/ PRIOR FILING DATE: 2001-08-08

/ PRIOR APPLICATION NUMBER: US 60/243,096

/ PRIOR FILING DATE: 2000-10-24

/ PRIOR APPLICATION NUMBER: US 60/252,147

/ PRIOR FILING DATE: 2000-11-20

/ PRIOR APPLICATION NUMBER: US 60/250,092

/ PRIOR FILING DATE: 2000-11-30

/ PRIOR APPLICATION NUMBER: US 60/261,766

/ PRIOR FILING DATE: 2001-01-16

/ PRIOR APPLICATION NUMBER: US 60/289,846

/ PRIOR FILING DATE: 2001-05-09

/ NUMBER OF SEQ ID NOS: 957086

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 887429

/ LENGTH: 621

/ TYPE: DNA

/ ORGANISM: Homo sapiens

US-09-925-065A-887429

Query Match 93.6%; Score 23.4; DB 4; Length 621;
Best Local Similarity 96.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||
Db 443 AAAAAAAAAATCACCACAAATCT 467

RESULT 16
US-09-925-065A-887430

/ Sequence 887430, Application US/09925065A
/ Publication No. US20050228172A9

/ GENERAL INFORMATION:

/ APPLICANT: Wang, David G.

/ TITLE OF INVENTION: Identification and Mapping of Single

/ FILE REFERENCE: 108827.135

/ CURRENT APPLICATION NUMBER: US/09/925,065A

/ PRIOR FILING DATE: 2001-08-08

/ PRIOR APPLICATION NUMBER: US 60/243,096

/ PRIOR FILING DATE: 2000-10-24

/ PRIOR APPLICATION NUMBER: US 60/252,147

/ PRIOR FILING DATE: 2000-11-20

/ PRIOR APPLICATION NUMBER: US 60/250,092

/ PRIOR FILING DATE: 2000-11-30

/ PRIOR APPLICATION NUMBER: US 60/261,766

/ PRIOR FILING DATE: 2001-01-16

/ PRIOR APPLICATION NUMBER: US 60/289,846

/ PRIOR FILING DATE: 2001-05-09

/ NUMBER OF SEQ ID NOS: 957086

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 887430

/ LENGTH: 621

/ TYPE: DNA

/ ORGANISM: Homo sapiens

US-09-925-065A-887430

Query Match 93.6%; Score 23.4; DB 4; Length 621;
Best Local Similarity 96.0%; Pred. No. 1.7e+02;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||
Db 443 AAAAAAAAAATCACCACAAATCT 467

RESULT 17

US-09-925-065A-831477/C

/ Sequence 831477, Application US/09925065A
/ Publication No. US20050228172A9

/ GENERAL INFORMATION:

/ APPLICANT: Wang, David G.

/ TITLE OF INVENTION: Identification and Mapping of Single

/ FILE REFERENCE: 108827.135

/ CURRENT APPLICATION NUMBER: US/09/925,065A

/ PRIOR FILING DATE: 2001-08-08

/ PRIOR APPLICATION NUMBER: US 60/243,096

/ PRIOR FILING DATE: 2000-10-24

/ PRIOR APPLICATION NUMBER: US 60/252,147

/ PRIOR FILING DATE: 2000-11-20

/ PRIOR APPLICATION NUMBER: US 60/250,092

/ PRIOR FILING DATE: 2000-11-30

/ PRIOR APPLICATION NUMBER: US 60/261,766

/ PRIOR FILING DATE: 2001-01-16

/ PRIOR APPLICATION NUMBER: US 60/289,846

/ PRIOR FILING DATE: 2001-05-09

/ NUMBER OF SEQ ID NOS: 957086

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 831477

```

; LENGTH: 689
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-831477

Query Match
Best Local Similarity 93.6%; Score 23.4; DB 4; Length 689;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 438 AAAAAAAAAATCGCAACCAATCT 414

RESULT 18
US-10-087-192-1504/C
; Sequence 1504, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1504
; LENGTH: 142318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(142318)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1504

Query Match
Best Local Similarity 93.6%; Score 23.4; DB 5; Length 142318;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 121397 AAAAAAAAAATCGCAACCAATCT 121373

RESULT 19
US-10-681-773-16
; Sequence 16, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 16
; LENGTH: 25
; TYPE: DNA
```

```

; ORGANISM: Homo sapien
US-10-681-773-16

Query Match
Best Local Similarity 92.0%; Score 23; DB 7; Length 25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 AAAAAAAAAATCGCAACCAATCT 25
Db 1 AAAAAAAAAATCGCAACCAATCT 23

RESULT 20
US-10-681-773-26
; Sequence 26, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 26
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-26

Query Match
Best Local Similarity 88.0%; Score 22; DB 7; Length 25;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AAAAAAAAAATCGCAACCAATCT 25
Db 1 AAAAAAAAAATCGCAACCAATCT 22

RESULT 21
US-10-741-600-17942/C
; Sequence 17942, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17942
; LENGTH: 58038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-17942

Query Match
Best Local Similarity 88.0%; Score 22; DB 8; Length 58038;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAAA 22
Db 51504 AAAAAAAAAATCGCAACCAAA 51483
```

```
RESULT 22
US-10-741-600-17681/c
; Sequence 17681, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01499
; CURRENT APPLICATION NUMBER: US/10/741,600
; NUMBER OF SEQ ID NOS: 2003-12-22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17681
; LENGTH: 333811
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(333811)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17681
```

```
Query Match      87.2%; Score 21.8; DB 5; Length 458;
Best Local Similarity 92.0%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 236397 AAAAAAAAAATCGCAACAA 236376
```

```
RESULT 23
US-10-027-632-57229
; Sequence 57229, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57229
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-57229
```

```
Query Match      87.2%; Score 21.8; DB 5; Length 458;
Best Local Similarity 92.0%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 128 AAAAAAAAAATCGCAACAAATCT 152
```

```
RESULT 24
US-10-027-632-57230
; Sequence 57230, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57230
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-57230
```

```
Query Match      87.2%; Score 21.8; DB 5; Length 458;
Best Local Similarity 92.0%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 128 AAAAAAAAAATCGCAACAAATCT 152
```

```
RESULT 25
US-10-027-632-309952
; Sequence 309952, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309952
; LENGTH: 458
```



```

; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309952

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 5; Length 458;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 128 AAAAAAAAAATCACCACAAAAAATCT 152

RESULT 26
US-10-027-632-309953
; Sequence 309953, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309953
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309953

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 5; Length 458;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 128 AAAAAAAAAATCACCACAAAAAATCT 152

RESULT 27
US-10-027-632-57229
; Sequence 57229, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
```

```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57229
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-57229

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 6; Length 458;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 128 AAAAAAAAAATCACCACAAAAAATCT 152

RESULT 28
US-10-027-632-57230
; Sequence 57230, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57230
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-57230

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 6; Length 458;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 128 AAAAAAAAAATCACCACAAAAAATCT 152

RESULT 29
US-10-027-632-309952
; Sequence 309952, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309952
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309952
```

```

Query Match      87.2%; Score 21.8; DB 6; Length 458;
Best Local Similarity 92.0%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAAATCCGCAACCAATCT 25
    |||||
Db 128 AAAAAAAAAATCCGCAACCAATCT 152
```

```

RESULT 30
US-10-027-632-309953
; Sequence 309953, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309953
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309953
```

```

Query Match      87.2%; Score 21.8; DB 6; Length 458;
Best Local Similarity 92.0%; Pred. No. 5.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAAATCCGCAACCAATCT 25
```

```

Db 128 AAAAAAAAAATCCGCAACCAATCT 152
```

```

RESULT 31
US-09-925-065A-384901
; Sequence 384901, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384901
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-384901
```

```

Query Match      87.2%; Score 21.8; DB 4; Length 539;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAAATCCGCAACCAATCT 25
    |||||
Db 508 AAAAAAAAAATCCGCAACCAATCT 532
```

```

RESULT 32
US-10-027-632-4931/C
; Sequence 4931, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4931
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-4931

Query Match 87.2%; Score 21.8; DB 5; Length 539;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||||
DB 508 AAAAAAAAAAGCAACAAATCT 484

RESULT 33

US-10-027-632-4931/C
Sequence 4931, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE OF INVENTION: Polymorphisms in the Human Genome

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4931

LENGTH: 539

TYPE: DNA

ORGANISM: Human

US-10-027-632-4931

Query Match 87.2%; Score 21.8; DB 6; Length 539;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||||
DB 508 AAAAAAAAAAGCAACAAATCT 484

RESULT 34

US-10-027-632-189916
Sequence 189916, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE OF INVENTION: Polymorphisms in the Human Genome

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 189916

LENGTH: 540

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(540)

OTHER INFORMATION: n = A,T,C or G

US-10-027-632-189916

Query Match 87.2%; Score 21.8; DB 5; Length 540;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||||
DB 37 AAAAAAAAAAGCAACAAATCT 61

RESULT 35

US-10-027-632-189917
Sequence 189917, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE OF INVENTION: Polymorphisms in the Human Genome

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 189917

LENGTH: 540

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(540)

OTHER INFORMATION: n = A,T,C or G

US-10-027-632-189917

Query Match 87.2%; Score 21.8; DB 5; Length 540;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||||
DB 37 AAAAAAAAAAGCAACAAATCT 61

RESULT 36

```
US-10-027-632-189916
; Sequence 189916, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 189916
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) ..(540)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-189916

Query Match      87.2% Score 21.8; DB 6; Length 540;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 37 AAAAAAAAAAGCAACCAATCT 61

RESULT 37
US-10-027-632-189917
; Sequence 189917, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 189917
```

```
US-10-027-632-189917
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) ..(540)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-189917

Query Match      87.2% Score 21.8; DB 6; Length 540;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 37 AAAAAAAAAAGCAACCAATCT 61

RESULT 38
US-09-925-065A-184309
; Sequence 184309, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 184309
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-184309

Query Match      87.2% Score 21.8; DB 4; Length 551;
Best Local Similarity 92.0%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 477 AAAAAAAAAATCGCAACCAATCT 501

RESULT 39
US-09-925-065A-702224
; Sequence 702224, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
```

```

; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 702224
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-702224

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 4; Length 590;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 108 AAAAAAATCACAACAAATCT 132

RESULT 40
US-09-925-065A-702225
; Sequence 702225, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 702225
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-702225

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 4; Length 590;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 108 AAAAAAATCACAACAAATCT 132

RESULT 41
US-09-925-065A-702226
; Sequence 702226, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
```

```

; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 702226
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-702226

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 4; Length 590;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 108 AAAAAAATCACAACAAATCT 132

RESULT 42
US-09-925-065A-726714
; Sequence 726714, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 726714
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-726714

Query Match
Best Local Similarity 87.2%; Score 21.8; DB 4; Length 636;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 216 AAAAAAATCGCAACAAATCT 240

RESULT 43
US-09-925-065A-726715
; Sequence 726715, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
```

```

; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 726715
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-925-065A-726715

Query Match      87.2% Score 21.8; DB 4; Length 636;
Best Local Similarity 92.0%; Pred. No. 6e+02; 2; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGCAACAATCT 25
    |||||
Db 216 AAAAAAAAAATCGCAACAATCT 240

RESULT 44
US-09-925-065A-726716
; Sequence 726716, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 726716
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-925-065A-726716

Query Match      87.2% Score 21.8; DB 4; Length 636;
Best Local Similarity 92.0%; Pred. No. 6e+02; 2; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGCAACAATCT 25
    |||||
Db 216 AAAAAAAAAATCGCAACAATCT 240

RESULT 45
US-10-027-632-70791
; Sequence 70791, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
```

```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70791
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Human
;
US-10-027-632-70791

Query Match      87.2% Score 21.8; DB 5; Length 640;
Best Local Similarity 92.0%; Pred. No. 6e+02; 2; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGCAACAATCT 25
    |||||
Db 90 AAAAAAAAAATCGCAACAATCT 114

RESULT 46
US-10-027-632-319807
; Sequence 319807, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319807
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Human
;
US-10-027-632-319807

Query Match      87.2% Score 21.8; DB 5; Length 640;
Best Local Similarity 92.0%; Pred. No. 6e+02; 2; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGCAACAATCT 25
    |||||
```

```
Db          90 AAAAAAAAAATTCACAAAAAATCT 114

RESULT 47
US-10-027-632-70791
; Sequence 70791, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70791
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70791.

Query Match          87.2%; Score 21.8; DB 6; Length 640;
Best Local Similarity 92.0%; Pred. No. 6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy          1 AAAAAAAAAATTCGACAAACAATCT 25
Db          90 AAAAAAAAAATTCACAAAAAATCT 114

RESULT 48
US-10-027-632-319807
; Sequence 319807, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319807
```

```
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-319807

Query Match          87.2%; Score 21.8; DB 6; Length 640;
Best Local Similarity 92.0%; Pred. No. 6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy          1 AAAAAAAAAATTCGACAAACAATCT 25
Db          411 AAAAAAAAAATTCACAAAAAATCT 387

RESULT 50
US-10-027-632-150512/c
; Sequence 150512, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29

RESULT 49
US-10-027-632-150512/c
; Sequence 150512, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150512
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-150512

Query Match          87.2%; Score 21.8; DB 5; Length 714;
Best Local Similarity 92.0%; Pred. No. 6.1e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy          1 AAAAAAAAAATTCGACAAACAATCT 25
Db          411 AAAAAAAAAATTCACAAAAAATCT 387
```

```
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 150512
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-150512
```

```
Query Match      87.2%; Score 21.8; DB 6; Length 714;
Best Local Similarity 92.0%; Pred. No. 6.1e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAATCGCAACAAATCT 25
         |||||||
Db      411 AAAAAAAAAATCACAATAATCT 387
```

Search completed: December 14, 2005, 08:46:05
Job time : 378.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 ; Search time 373.2 Seconds
(without alignments)
553.951 Million cell updates/sec

Title: US-10-681-773-2

Perfect score: 25
Sequence: 1 aaaaaaaaaacgcagacaacatc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Published Applications_NA_Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	7	US-10-681-773-2
2	24.6	98.4	33	9	US-10-891-260-1793
3	24.6	98.4	552	5	US-10-027-632-78265
4	24.6	98.4	552	5	US-10-027-632-78266
5	24.6	98.4	552	5	US-10-027-632-300721
6	24.6	98.4	552	5	US-10-027-632-300722
7	24.6	98.4	552	6	US-10-027-632-78265
8	24.6	98.4	552	6	US-10-027-632-78266
9	24.6	98.4	552	6	US-10-027-632-300721
10	24.6	98.4	552	6	US-10-027-632-300722
11	23.4	93.6	25	7	US-10-681-773-1
12	23.4	93.6	25	7	US-10-719-993-7034
13	23	92.0	25	7	US-10-681-773-17
14	22	88.0	25	7	US-10-681-773-27
15	21.8	87.2	621	4	US-09-925-065A-887427
16	21.8	87.2	621	4	US-09-925-065A-887428
17	21.8	87.2	621	4	US-09-925-065A-887429
18	21.8	87.2	621	4	US-09-925-065A-887430
19	21.8	87.2	689	4	US-09-925-065A-831477
20	21.8	87.2	142318	5	US-10-087-192-1504
21	21.4	85.6	25	7	US-10-681-773-16
22	21	84.0	25	7	US-10-681-773-58
23	20.8	83.2	2300	4	US-09-925-065A-725864

C	24	20.8	83.2	3673778	6	US-10-312-841-2	Sequence 2, Appli
C	25	20.4	81.6	25	7	US-10-681-773-26	Sequence 26, Appli
C	26	20.4	81.6	442	5	US-10-027-632-201656	Sequence 201656,
C	27	20.4	81.6	442	6	US-10-027-632-201656	Sequence 201656,
C	28	20.4	81.6	1988	4	US-09-925-065A-677554	Sequence 677554,
C	29	20.4	81.6	1988	4	US-09-925-065A-677555	Sequence 677555,
C	30	20.4	81.6	1988	4	US-09-925-065A-677556	Sequence 677556,
C	31	20.4	81.6	58038	8	US-10-741-600-17942	Sequence 17942, A
C	32	20.4	81.6	333811	8	US-10-741-600-17681	Sequence 17681, A
C	33	20.2	80.8	367	5	US-09-925-065A-152895	Sequence 152895,
C	34	20.2	80.8	458	5	US-10-027-632-57229	Sequence 57229, A
C	35	20.2	80.8	458	5	US-10-027-632-57230	Sequence 57230, A
C	36	20.2	80.8	458	5	US-10-027-632-309952	Sequence 309952,
C	37	20.2	80.8	458	5	US-10-027-632-309953	Sequence 309953,
C	38	20.2	80.8	458	6	US-10-027-632-57229	Sequence 57229, A
C	39	20.2	80.8	458	6	US-10-027-632-57230	Sequence 57230, A
C	40	20.2	80.8	458	6	US-10-027-632-309952	Sequence 309952,
C	41	20.2	80.8	458	6	US-10-027-632-309953	Sequence 309953,
C	42	20.2	80.8	488	4	US-09-925-065A-152894	Sequence 152894,
C	43	20.2	80.8	539	4	US-09-925-065A-384901	Sequence 384901,
C	44	20.2	80.8	539	5	US-10-027-632-4931	Sequence 4931, Ap
C	45	20.2	80.8	539	5	US-10-027-632-4931	Sequence 4931, Ap
C	46	20.2	80.8	540	6	US-10-027-632-189916	Sequence 189916,
C	47	20.2	80.8	540	6	US-10-027-632-189917	Sequence 189917,
C	48	20.2	80.8	540	6	US-10-027-632-189916	Sequence 189916,
C	49	20.2	80.8	540	6	US-10-027-632-189917	Sequence 189917,
C	50	20.2	80.8	551	4	US-09-925-065A-184309	Sequence 184309,
C	51	20.2	80.8	590	4	US-09-925-065A-66291	Sequence 66291, A
C	52	20.2	80.8	590	4	US-09-925-065A-702224	Sequence 702224,
C	53	20.2	80.8	590	4	US-09-925-065A-702225	Sequence 702225,
C	54	20.2	80.8	590	4	US-09-925-065A-702226	Sequence 702226,
C	55	20.2	80.8	607	4	US-09-925-065A-780136	Sequence 780136,
C	56	20.2	80.8	607	4	US-09-925-065A-780137	Sequence 780137,
C	57	20.2	80.8	636	4	US-09-925-065A-726714	Sequence 726714,
C	58	20.2	80.8	636	4	US-09-925-065A-726715	Sequence 726715,
C	59	20.2	80.8	636	4	US-09-925-065A-726716	Sequence 726716,
C	60	20.2	80.8	640	5	US-10-027-632-70791	Sequence 70791, A
C	61	20.2	80.8	640	5	US-10-027-632-319807	Sequence 319807,
C	62	20.2	80.8	640	6	US-10-027-632-319807	Sequence 319807,
C	63	20.2	80.8	640	6	US-10-027-632-319807	Sequence 319807,
C	64	20.2	80.8	714	6	US-10-027-632-150512	Sequence 150512,
C	65	20.2	80.8	714	6	US-10-027-632-150512	Sequence 150512,
C	66	20.2	80.8	781	7	US-10-437-963-72628	Sequence 72628, A
C	67	20.2	80.8	840	4	US-09-925-065A-691495	Sequence 691495,
C	68	20.2	80.8	840	4	US-09-925-065A-691496	Sequence 691496,
C	69	20.2	80.8	840	4	US-09-925-065A-691497	Sequence 691497,
C	70	20.2	80.8	864	5	US-10-027-632-38334	Sequence 38334, A
C	71	20.2	80.8	864	6	US-10-027-632-38334	Sequence 38334, A
C	72	20.2	80.8	902	5	US-10-027-632-4698	Sequence 4698, Ap
C	73	20.2	80.8	902	5	US-10-027-632-4698	Sequence 4698, Ap
C	74	20.2	80.8	902	6	US-10-027-632-4698	Sequence 4698, Ap
C	75	20.2	80.8	902	6	US-10-027-632-4698	Sequence 4698, Ap
C	76	20.2	80.8	989	7	US-10-424-599-122678	Sequence 122678,
C	77	20.2	80.8	1410	4	US-09-925-065A-547583	Sequence 547583,
C	78	20.2	80.8	3875	10	US-11-097-143-07076	Sequence 40726, A
C	79	20.2	80.8	6162	7	US-10-240-5899C-124	Sequence 124, App
C	80	20.2	80.8	6226	3	US-09-764-847-1961	Sequence 1961, App
C	81	20.2	80.8	6226	3	US-10-027-632-4698	Sequence 4698, Ap
C	82	20.2	80.8	7145	7	US-10-027-632-4698	Sequence 4698, Ap
C	83	20.2	80.8	7928	6	US-10-311-455-68	Sequence 43, Appl
C	84	20.2	80.8	7928	6	US-10-311-455-68	Sequence 68, Appl
C	85	20.2	80.8	29105	6	US-10-419-723-1	Sequence 12, Appl
C	86	20.2	80.8	41936	9	US-09-967-7681A-116	Sequence 116, App
C	87	20.2	80.8	41936	9	US-10-843-661A-1261	Sequence 6261, Ap
C	88	20.2	80.8	159759	7	US-10-322-281-10	Sequence 10, Appl
C	89	20.2	80.8	159759	7	US-10-322-281-10	Sequence 308, App
C	90	20.2	80.8	1691139	5	US-10-067-514-1	Sequence 1, Appli
C	91	20.2	80.8	1691139	7	US-10-419-723-1	Sequence 1, Appli
C	92	20.2	80.8	1691139	7	US-10-255-120-1	Sequence 1, Appli
C	93	20.2	80.8	1691139	9	US-10-868-197-1	Sequence 158, Appl
C	94	20	80.0	25	7	US-10-681-773-158	Sequence 324, App
C	95	19.8	79.2	230	3	US-09-854-667-324	Sequence 324, App
C	96	19.8	79.2	230	9	US-10-786-970A-324	Sequence 324, App

```
97 19.8 79.2 267 7 US-10-021-323-5918 Sequence 5918, Ap
98 19.8 79.2 435 7 US-10-424-599-27981 Sequence 27981, A
99 19.8 79.2 539 4 US-09-925-065A-384902 Sequence 384902,
C 100 19.8 79.2 577 8 US-10-363-345A-8501 Sequence 8501, Ap
C 101 19.8 79.2 577 8 US-10-363-345A-8502 Sequence 8502, Ap
C 102 19.8 79.2 577 9 US-10-363-483A-8501 Sequence 8501, Ap
C 103 19.8 79.2 577 9 US-10-363-483A-8502 Sequence 8502, Ap
C 104 19.8 79.2 609 4 US-09-925-065A-641478 Sequence 641478, Ap
C 105 19.8 79.2 1710 6 US-10-311-455-2258 Sequence 2258, Ap
C 106 19.8 79.2 5596 3 US-09-728-552-5 Sequence 5, Appl1
C 107 19.8 79.2 52985 3 US-10-756-149-4085 Sequence 4085, Ap
C 108 19.8 79.2 80246 3 US-09-728-552-4 Sequence 4, Appl1
C 109 19.8 79.2 80595 3 US-09-728-552-3 Sequence 3, Appl1
C 110 19.8 79.2 379652 9 US-10-481-613-71 Sequence 71, Appl
C 111 19.8 79.2 25 5 US-10-681-773-57 Sequence 57, Appl
C 112 19.4 77.6 751 5 US-10-027-632-15723 Sequence 15723, A
C 113 19.4 77.6 751 6 US-10-027-632-15723 Sequence 15723, A
C 114 19.4 77.6 1149 4 US-09-925-065A-689860 Sequence 689860,
C 115 19.2 76.8 413 7 US-10-663-561-415 Sequence 415, App
C 116 19.2 76.8 422 4 US-09-925-065A-351562 Sequence 351562,
C 117 19.2 76.8 498 7 US-10-724-972A-2237 Sequence 2237, Ap
C 118 19.2 76.8 544 8 US-09-925-065A-110877 Sequence 110877,
C 119 19.2 76.8 544 8 US-10-363-345A-12417 Sequence 12417, A
C 120 19.2 76.8 544 8 US-10-363-345A-12418 Sequence 12418, A
C 121 19.2 76.8 544 9 US-10-363-483A-12417 Sequence 12417, A
C 122 19.2 76.8 544 9 US-10-363-483A-12418 Sequence 12418, A
C 123 19.2 76.8 545 4 US-09-925-065A-93633 Sequence 93633, A
C 124 19.2 76.8 555 4 US-09-925-065A-108610 Sequence 108610,
C 125 19.2 76.8 570 4 US-09-925-065A-254576 Sequence 254576,
C 126 19.2 76.8 570 4 US-09-925-065A-254577 Sequence 254577,
C 127 19.2 76.8 570 4 US-09-925-065A-254578 Sequence 254578,
C 128 19.2 76.8 587 8 US-10-425-115-93990 Sequence 93990, A
C 129 19.2 76.8 594 4 US-09-925-065A-718544 Sequence 718544,
C 130 19.2 76.8 600 9 US-10-972-079-31015 Sequence 31015, A
C 131 19.2 76.8 600 9 US-10-972-079-31016 Sequence 31016, A
C 132 19.2 76.8 612 8 US-10-363-345A-11035 Sequence 11035, A
C 133 19.2 76.8 612 8 US-10-363-345A-11036 Sequence 11036, A
C 134 19.2 76.8 612 9 US-10-363-483A-11035 Sequence 11035, A
C 135 19.2 76.8 612 9 US-10-363-483A-11036 Sequence 11036, A
C 136 19.2 76.8 630 5 US-10-027-632-139935 Sequence 139935,
C 137 19.2 76.8 630 6 US-10-027-632-139935 Sequence 139935,
C 138 19.2 76.8 635 4 US-09-925-065A-699791 Sequence 699791,
C 139 19.2 76.8 635 4 US-09-925-065A-699792 Sequence 699792,
C 140 19.2 76.8 635 4 US-09-925-065A-699793 Sequence 699793,
C 141 19.2 76.8 639 4 US-09-925-065A-715390 Sequence 715390,
C 142 19.2 76.8 643 4 US-09-925-065A-451399 Sequence 451399,
C 143 19.2 76.8 672 4 US-09-925-065A-63495 Sequence 63495, A
C 144 19.2 76.8 722 8 US-10-363-345A-21751 Sequence 21751, A
C 145 19.2 76.8 722 8 US-10-363-345A-21752 Sequence 21752, A
C 146 19.2 76.8 722 9 US-10-363-483A-21751 Sequence 21751, A
C 147 19.2 76.8 722 9 US-10-363-483A-21752 Sequence 21752, A
C 148 19.2 76.8 730 8 US-10-363-345A-1811 Sequence 1811, Ap
C 149 19.2 76.8 730 8 US-10-363-345A-1812 Sequence 1812, Ap
C 150 19.2 76.8 730 9 US-10-363-483A-1811 Sequence 1811, Ap
```

ALIGNMENTS

```
RESULT 1
US-10-681-773-2
; Sequence 2, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 2
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-2
```

```
Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGACGACCAATCT 25
Db 1 AAAAAAAAAATCGACGACCAATCT 25
```

```
RESULT 2
US-10-891-260-1793
; Sequence 1793, Application US/10891260
; Publication No. US20050227244A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Affymetrix, Inc.
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
```

```
; TITLE OF INVENTION: Methods of Analysis of Human Polymorphisms
```

```
; FILE REFERENCE: 3522.3
```

```
; CURRENT APPLICATION NUMBER: US/10/891,260
```

```
; PRIOR FILING DATE: 2004-07-13
```

```
; PRIOR APPLICATION NUMBER: 10/681,773
```

```
; NUMBER OF SEQ ID NOS: 10244
```

```
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 1793
```

```
; LENGTH: 33
```

```
; TYPE: DNA
```

```
; ORGANISM: homo sapien
```

```
US-10-891-260-1793
Query Match 98.4%; Score 24.6; DB 9; Length 33;
Best Local Similarity 96.0%; Pred. No. 17;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGACGACCAATCT 25
Db 1 AAAAAAAAAATCGACGACCAATCT 25
```

```
RESULT 3
US-10-027-632-78265
; Sequence 78265, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```
; FILE REFERENCE: 108827.129
```

```
; CURRENT APPLICATION NUMBER: US/10/027,632
```

```
; PRIOR FILING DATE: 2002-04-30
```

```
; PRIOR APPLICATION NUMBER: US 60/218,006
```

```
; PRIOR FILING DATE: 2000-07-12
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; PRIOR FILING DATE: 2000-04-20
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
```

```
; PRIOR FILING DATE: 2000-03-29
```

```
; PRIOR APPLICATION NUMBER: US 60/185,218
```

```
; PRIOR FILING DATE: 2000-02-24
```

```
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```
; PRIOR FILING DATE: 1999-11-23
```

```
; PRIOR APPLICATION NUMBER: US 60/156,358
```

```

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78265
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-78265

Query Match          98.4%; Score 24.6; DB 5; Length 552;
Best Local Similarity 96.0%; Pred. No. 22;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 442 AAAAAAAAAATCGCAGCAAAATCT 466

RESULT 4
US-10-027-632-78266
; Sequence 78266, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78266
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-78266

Query Match          98.4%; Score 24.6; DB 5; Length 552;
Best Local Similarity 96.0%; Pred. No. 22;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 442 AAAAAAAAAATCGCAGCAAAATCT 466

RESULT 5
US-10-027-632-300721
; Sequence 300721, Application US/10027632
```

```

; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300721
; LENGTH: 552
; TYPE: DNA

Query Match          98.4%; Score 24.6; DB 5; Length 552;
Best Local Similarity 96.0%; Pred. No. 22;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 442 AAAAAAAAAATCGCAGCAAAATCT 466

RESULT 6
US-10-027-632-300722
; Sequence 300722, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300722
; LENGTH: 552
; TYPE: DNA
```

```

; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-300722
```

```

Query Match          98.4%; Score 24.6; DB 5; Length 552;
Best Local Similarity 96.0%; Pred. No. 22;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

Oy 1 AAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 442 AAAAAAAAAATCGCAGCAAAATCT 466
```

RESULT 7

```

US-10-027-632-78265
; Sequence 78265, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78265
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-78265
```

```

Query Match          98.4%; Score 24.6; DB 6; Length 552;
Best Local Similarity 96.0%; Pred. No. 22;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

Oy 1 AAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 442 AAAAAAAAAATCGCAGCAAAATCT 466
```

RESULT 8

```

US-10-027-632-78266
; Sequence 78266, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-300721
```

```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78266
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-78266
```

```

Query Match          98.4%; Score 24.6; DB 6; Length 552;
Best Local Similarity 96.0%; Pred. No. 22;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

Oy 1 AAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 442 AAAAAAAAAATCGCAGCAAAATCT 466
```

RESULT 9

```

US-10-027-632-300721
; Sequence 300721, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300721
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(552)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-300721
```

```

Query Match          98.4%; Score 24.6; DB 6; Length 552;
```

Best Local Similarity 96.0%; Pred. No. 22;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGACAAATCT 25
|||||
Db 442 AAAAAAAAAATCGCAGACAAATCT 466

RESULT 10
US-10-027-632-300722

; Sequence 300722, Application US/10027632
; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 300722

; LENGTH: 552

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1) - (552)

; OTHER INFORMATION: n = A,T,C or G

US-10-027-632-300722

Query Match 98.4%; Score 24.6; DB 6; Length 552;
Best Local Similarity 96.0%; Pred. No. 22;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGACAAATCT 25
|||||
Db 442 AAAAAAAAAATCGCAGACAAATCT 466

RESULT 11
US-10-681-773-1

; Sequence 1, Application US/10681773
; Publication No. US20040146890A1

; GENERAL INFORMATION:

; APPLICANT: Matsuzaki, Hajime

; APPLICANT: Wei, Rui

; APPLICANT: Shen, Mei-Mei

; APPLICANT: Kennedy, Giulia

; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans

; FILE REFERENCE: 3522.2

; CURRENT APPLICATION NUMBER: US/10/681,773

; CURRENT FILING DATE: 2003-10-07

; PRIOR APPLICATION NUMBER: 60/470,475

; PRIOR FILING DATE: 2002-05-14

; PRIOR APPLICATION NUMBER: 60/417,190

; PRIOR FILING DATE: 2002-10-08

; NUMBER OF SEQ ID NOS: 124031

; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1

; SEQ ID NO 1
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-1

Query Match 93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 44;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGACAAATCT 25
|||||
Db 1 AAAAAAAAAATCGCAGACAAATCT 25

RESULT 12

US-10-719-993-7034

; Sequence 7034, Application US/10719993
; Publication No. US20040265849A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; FILE REFERENCE: CLO01496

; CURRENT APPLICATION NUMBER: US/10/719,993

; CURRENT FILING DATE: 2003-11-24

; NUMBER OF SEQ ID NOS: 55342

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7034

; LENGTH: 25473

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-719-993-7034

Query Match 93.6%; Score 23.4; DB 8; Length 25473;
Best Local Similarity 96.0%; Pred. No. 92;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGACAAATCT 25
|||||
Db 2923 AAAAAAAAAATCGCAGACAAATCT 2947

RESULT 13

US-10-681-773-17

; Sequence 17, Application US/10681773
; Publication No. US20040146890A1

; GENERAL INFORMATION:

; APPLICANT: Matsuzaki, Hajime

; APPLICANT: Wei, Rui

; APPLICANT: Shen, Mei-Mei

; APPLICANT: Kennedy, Giulia

; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans

; FILE REFERENCE: 3522.2

; CURRENT APPLICATION NUMBER: US/10/681,773

; CURRENT FILING DATE: 2003-10-07

; PRIOR APPLICATION NUMBER: 60/470,475

; PRIOR FILING DATE: 2002-05-14

; PRIOR APPLICATION NUMBER: 60/417,190

; PRIOR FILING DATE: 2002-10-08

; NUMBER OF SEQ ID NOS: 124031

; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1

; SEQ ID NO 17

; LENGTH: 25

; TYPE: DNA

; ORGANISM: Homo sapien

US-10-681-773-17

Query Match 92.0%; Score 23; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAAAAAAAAATCGCAGACAAATCT 25

Db 1 |||||
1 AAAAAAAAAATCGACGACCAATCT 23

RESULT 14
US-10-681-773-27
; Sequence 27, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 352.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 27
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-27

Query Match 88.0%; Score 22; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 AAAAAAAAAATCGACGACCAATCT 25
|||
Db 1 AAAAAAAAAATCGACGACCAATCT 22

RESULT 15
US-09-925-065A-887427
; Sequence 887427, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887427
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-887427

Query Match 87.2%; Score 21.8; DB 4; Length 621;
Best Local Similarity 92.0%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACGACCAATCT 25
|||

Db 443 AAAAAAAAAATCGACGACCAATCT 467

RESULT 16
US-09-925-065A-887428
; Sequence 887428, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887428
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-887428

Query Match 87.2%; Score 21.8; DB 4; Length 621;
Best Local Similarity 92.0%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACGACCAATCT 25
|||
Db 443 AAAAAAAAAATCGACGACCAATCT 467

RESULT 17
US-09-925-065A-887429
; Sequence 887429, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887429
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-887429

Query Match 87.2%; Score 21.8; DB 4; Length 621;
Best Local Similarity 92.0%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGGACCAATCT 25
|||||
Db 443 AAAAAAAAAATCGGACCAATCT 467

RESULT 18

US-09-925-065A-887430
; Sequence 887430, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887430
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-887430

Query Match 87.2%; Score 21.8; DB 4; Length 621;
Best Local Similarity 92.0%; Pred. No. 2.3e+02;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGGACCAATCT 25
|||||
Db 443 AAAAAAAAAATCGGACCAATCT 467

RESULT 19

US-09-925-065A-831477/c
; Sequence 831477, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 831477
; LENGTH: 689
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-831477

Query Match 87.2%; Score 21.8; DB 4; Length 689;
Best Local Similarity 92.0%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGGACCAATCT 25
|||||
Db 438 AAAAAAAAAATCGGACCAATCT 414

RESULT 20

US-10-087-192-1504/c
; Sequence 1504, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1504
; LENGTH: 142318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(142318)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1504

Query Match 87.2%; Score 21.8; DB 5; Length 142318;
Best Local Similarity 92.0%; Pred. No. 4.2e+02;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGGACCAATCT 25
|||||
Db 121397 AAAAAAAAAATCGGACCAATCT 121373

RESULT 21

US-10-681-773-16
; Sequence 16, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 16
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-16

Query Match 85.6%; Score 21.4; DB 7; Length 25;
Best Local Similarity 95.7%; Pred. No. 2.3e+02;

Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAAAAAAAAATCGCAGCAAAATCT 25
Db 1 AAAAAAAAAATCGCAGCAAAATCT 23

RESULT 22
US-10-681-773-58

/ Sequence 58, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsumaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ PRIOR FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO: 58
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-58

Query Match
Best Local Similarity 100.0%; Score 21; DB 7; Length 25;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AAAAAAAAAATCGCAGCAAAATCT 25
Db 1 AAAAAAAAAATCGCAGCAAAATCT 21

RESULT 23

US-09-925-065A-725864/C
/ Sequence 725864, Application US/09925065A
/ Publication No. US2005028172A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ FILE REFERENCE: 108827.135
/ CURRENT APPLICATION NUMBER: US/09/925,065A
/ PRIOR FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243,096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289,846
/ PRIOR FILING DATE: 2001-05-09
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 725864
/ LENGTH: 2300
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-925-065A-725864

Query Match
Best Local Similarity 91.7%; Score 20.8; DB 4; Length 2300;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGCAGCAAAATCT 25
Db 1593 AAAAAAAAAAGGCGAGCAAAATCT 1570

RESULT 24

US-10-312-841-2/C
/ Sequence 2, Application US/10312841
/ Publication No. US20030186277A1
/ GENERAL INFORMATION:
/ APPLICANT: Epigenomics AG
/ TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
/ FILE REFERENCE: E01/1208/WO
/ CURRENT APPLICATION NUMBER: US/10/312,841
/ PRIOR FILING DATE: 2002-12-30
/ NUMBER OF SEQ ID NOS: 2
/ SEQ ID NO: 2
/ LENGTH: 3673778
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (379615)
US-10-312-841-2

Query Match
Best Local Similarity 91.7%; Score 20.8; DB 6; Length 3673778;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATC 24
Db 3517562 AAAAAAAAAATCGCAGCAAAATC 3517539

RESULT 25

US-10-681-773-26
/ Sequence 26, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsumaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ PRIOR FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO: 26
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-26

Query Match
Best Local Similarity 95.5%; Score 20.4; DB 7; Length 25;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 AAAAAAAAAATCGCAGCAAAATCT 25
Db 1 AAAAAAAAAATCGCAGCAAAATCT 22

RESULT 26
US-10-027-632-201656/C


```
Sequence 201656, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 201656
LENGTH: 442
TYPE: DNA
ORGANISM: Human
US-10-027-632-201656
```

```
Query Match      81.6%; Score 20.4; DB 5; Length 442;
Best Local Similarity 87.5%; Pred. No. 7.2e+02;
Matches 21; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 AAAAAAAAAATCGACAGCAATCT 25
Db 345 AAAAAAAAAATCGACAGCAATCT 322
```

```
RESULT 27
US-10-027-632-201656/c
Sequence 201656, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 201656
LENGTH: 442
TYPE: DNA
ORGANISM: Human
US-10-027-632-201656
```

```
Query Match      81.6%; Score 20.4; DB 6; Length 442;
Best Local Similarity 95.5%; Pred. No. 8.5e+02;
Matches 21; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 AAAAAAAAAATCGACAGCAATCT 25
Db 345 AAAAAAAAAATCGACAGCAATCT 322
```

```
RESULT 28
US-09-925-065A-677554
Sequence 677554, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 677554
LENGTH: 1988
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-677554
```

```
Query Match      81.6%; Score 20.4; DB 4; Length 1988;
Best Local Similarity 95.5%; Pred. No. 8.5e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGACAGCAAA 22
Db 1192 AAAAAAAAAATCGACAGCAAA 1213
```

```
RESULT 29
US-09-925-065A-677555
Sequence 677555, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 677555
LENGTH: 1988
TYPE: DNA
```

```

; ORGANISM: Homo sapiens
US-09-925-065A-677555

Query Match      81.6%; Score 20.4; DB 4; Length 1988;
Best Local Similarity 95.5%; Pred. No. 8.5e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACAGCAAA 22
Db 1192 AAAAAAAAAATCGACAGCAAA 1213

RESULT 30
US-09-925-065A-677556
; Sequence 677556, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 677556
; LENGTH: 1988
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-677556

Query Match      81.6%; Score 20.4; DB 4; Length 1988;
Best Local Similarity 95.5%; Pred. No. 8.5e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACAGCAAA 22
Db 1192 AAAAAAAAAATCGACAGCAAA 1213

RESULT 31
US-10-741-600-17942/C
; Sequence 17942, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; PRIOR FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17942
; LENGTH: 58038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-17942

Query Match      81.6%; Score 20.4; DB 8; Length 58038;
Best Local Similarity 95.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 AAAAAAAAAATCGACAGCAAA 22
Db 51504 AAAAAAAAAATCGACAGCAAA 51483

RESULT 32
US-10-741-600-17681/C
; Sequence 17681, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; PRIOR FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17681
; LENGTH: 333811
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(333811)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17681

Query Match      81.6%; Score 20.4; DB 8; Length 333811;
Best Local Similarity 95.5%; Pred. No. 1.4e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACAGCAAA 22
Db 236397 AAAAAAAAAATCGACAGCAAA 236376

RESULT 33
US-09-925-065A-152895
; Sequence 152895, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152895
; LENGTH: 367
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-152895

Query Match      80.8%; Score 20.2; DB 4; Length 367;
Best Local Similarity 88.0%; Pred. No. 8.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACAGCAAAATCT 25
Db 154 AAAAAAAAAATCGACAGCAAAATCT 178
```

```
RESULT 34
US-10-027-632-57229
; Sequence 57229, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57229
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-57229

Query Match      80.8%; Score 20.2; DB 5; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGCAGACAATCT 25
    |||||
Db 128 AAAAAAAAAATCGCAGACAATCT 152

RESULT 35
US-10-027-632-57230
; Sequence 57230, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57230
; LENGTH: 458
```

```
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-57230

Query Match      80.8%; Score 20.2; DB 5; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGCAGACAATCT 25
    |||||
Db 128 AAAAAAAAAATCGCAGACAATCT 152

RESULT 36
US-10-027-632-309952
; Sequence 309952, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309952
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309952

Query Match      80.8%; Score 20.2; DB 5; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGCAGACAATCT 25
    |||||
Db 128 AAAAAAAAAATCGCAGACAATCT 152

RESULT 37
US-10-027-632-309953
; Sequence 309953, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
```

```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309953
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309953

Query Match      80.8%; Score 20.2; DB 5; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 128 AAAAAAAAAATCGAGCAAAATCT 152

RESULT 38
US-10-027-632-57229
; Sequence 57229, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57229
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-57229

Query Match      80.8%; Score 20.2; DB 6; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 128 AAAAAAAAAATCGAGCAAAATCT 152

RESULT 39
US-10-027-632-57230
; Sequence 57230, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309952
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309952

Query Match      80.8%; Score 20.2; DB 6; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 128 AAAAAAAAAATCGAGCAAAATCT 152

RESULT 40
US-10-027-632-309952
; Sequence 309952, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309952
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309952

Query Match      80.8%; Score 20.2; DB 6; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Db      128 AAAAAAAAAATCAGCAAAATCT 152

RESULT 41
US-10-027-632-309953
; Sequence 309953, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 309953
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-309953

Query Match      80.8%; Score 20.2; DB 6; Length 458;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGAGCAAAATCT 25
Db      128 AAAAAAAAAATCAGCAAAATCT 152

RESULT 42
US-09-925-065A-152894
; Sequence 152894, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152894
; LENGTH: 488
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-09-925-065A-152894

Query Match      80.8%; Score 20.2; DB 4; Length 488;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGAGCAAAATCT 25
Db      420 AAAAAAAAAATTGAAGCAAAATGT 444

RESULT 43
US-09-925-065A-384901
; Sequence 384901, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384901
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-384901

Query Match      80.8%; Score 20.2; DB 4; Length 539;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGAGCAAAATCT 25
Db      508 AAAAAAAAAAGCAAAATCT 532

RESULT 44
US-10-027-632-4931/C
; Sequence 4931, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
```

```

; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4931
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-4931

Query Match      80.8%; Score 20.2; DB 5; Length 539;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 508 AAAAAAAAAAGCAAAATCT 484

RESULT 45
US-10-027-632-4931/C
; Sequence 4931, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4931
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-4931

Query Match      80.8%; Score 20.2; DB 6; Length 539;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 508 AAAAAAAAAAGCAAAATCT 484

RESULT 46
US-10-027-632-189916
; Sequence 189916, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
```

```

; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 189916
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(540)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-189917

Query Match      80.8%; Score 20.2; DB 5; Length 540;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 37 AAAAAAAAAAGCAAAATCT 61

RESULT 47
US-10-027-632-189917
; Sequence 189917, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 189917
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(540)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-189917

Query Match      80.8%; Score 20.2; DB 5; Length 540;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGACGACAAATCT 25
    |||||
    37 AAAAAAAAAAGCAACAAATCT 61

RESULT 48
US-10-027-632-189916
; Sequence 189916, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 189916
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(540)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-189916

Query Match      80.8%; Score 20.2; DB 6; Length 540;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGACAAATCT 25
    |||||
    37 AAAAAAAAAAGCAACAAATCT 61

RESULT 49
US-10-027-632-189917
; Sequence 189917, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
```

```
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 189917
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(540)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-189917

Query Match      80.8%; Score 20.2; DB 6; Length 540;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGACAAATCT 25
    |||||
    37 AAAAAAAAAAGCAACAAATCT 61

RESULT 50
US-09-925-065A-184309
; Sequence 184309, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 184309
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-184309

Query Match      80.8%; Score 20.2; DB 4; Length 551;
Best Local Similarity 88.0%; Pred. No. 8.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGACAAATCT 25
    |||||
    477 AAAAAAAAAATCGACGACAAATCT 501
```

Search completed: December 14, 2005, 08:46:15
Job time : 383.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model1

Run on: December 14, 2005, 02:03:56 ; Search time 373.2 Seconds

(without alignments)
553.951 Million cell updates/sec

Title: US-10-681-773-3

Perfect score: 25

Sequence: 1 aaaaaaaaaactaaagcttgatcctt 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :
Published Applications NA Main: *
1: /cgn2_6/ptodata/1/pubphna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/1/pubphna/US08_PUBCOMB.seq: *
3: /cgn2_6/ptodata/1/pubphna/US09_PUBCOMB.seq: *
4: /cgn2_6/ptodata/1/pubphna/US10_PUBCOMB.seq: *
5: /cgn2_6/ptodata/1/pubphna/US10A_PUBCOMB.seq: *
6: /cgn2_6/ptodata/1/pubphna/US10C_PUBCOMB.seq: *
7: /cgn2_6/ptodata/1/pubphna/US10D_PUBCOMB.seq: *
8: /cgn2_6/ptodata/1/pubphna/US10E_PUBCOMB.seq: *
9: /cgn2_6/ptodata/1/pubphna/US10F_PUBCOMB.seq: *
10: /cgn2_6/ptodata/1/pubphna/US11_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	US-10-681-773-3	Sequence 3, Appl1
2	24.6	98.4	525	US-10-027-632-63078	Sequence 63078, A
3	24.6	98.4	525	US-10-027-632-63079	Sequence 63079, A
4	24.6	98.4	525	US-10-027-632-63080	Sequence 63080, A
5	24.6	98.4	525	US-10-027-632-63081	Sequence 63081, A
6	24.6	98.4	525	US-10-027-632-63078	Sequence 63078, A
7	24.6	98.4	525	US-10-027-632-63079	Sequence 63079, A
8	24.6	98.4	525	US-10-027-632-63080	Sequence 63080, A
9	24.6	98.4	525	US-10-027-632-63081	Sequence 63081, A
10	24	96.0	25	US-10-681-773-10	Sequence 10, Appl1
11	24	96.0	25	US-10-681-773-87004	Sequence 87004, A
12	23.6	94.4	501	US-10-027-632-39077	Sequence 39077, A
13	23.6	94.4	501	US-10-027-632-39078	Sequence 39078, A
14	23.6	94.4	501	US-10-027-632-39079	Sequence 39079, A
15	23.6	94.4	501	US-10-027-632-39080	Sequence 39080, A
16	23.6	94.4	501	US-10-027-632-39077	Sequence 39077, A
17	23.6	94.4	501	US-10-027-632-39078	Sequence 39078, A
18	23.6	94.4	501	US-10-027-632-39079	Sequence 39079, A
19	23.6	94.4	501	US-10-027-632-39080	Sequence 39080, A
20	23.6	94.4	501	US-10-027-632-39077	Sequence 39077, A
21	23	92.0	25	US-10-681-773-4	Sequence 4, Appl1
22	22.4	89.6	25	US-10-681-773-11	Sequence 11, Appl1
23	22.4	89.6	25	US-10-681-773-87005	Sequence 87005, A

C 24	22.4	89.6	628	4	US-09-925-065A-800109
C 25	22.4	89.6	628	4	US-09-925-065A-800110
C 26	22.4	89.6	628	4	US-09-925-065A-800111
C 27	22.4	89.6	628	4	US-09-925-065A-854017
C 28	21.4	85.6	25	7	US-10-681-773-21
C 29	21.4	85.6	1666	7	US-10-425-114-14574
C 30	20.8	83.2	521	4	US-09-925-065A-133939
C 31	20.8	83.2	13606	5	US-10-311-676-166
C 32	20.8	83.2	13606	6	US-10-311-455-1784
C 33	20.8	83.2	13606	6	US-10-240-453-188
C 34	20.8	83.2	13606	7	US-10-221-1714-284
C 35	20.2	80.8	424	8	US-10-357-930-4072
C 36	20.2	80.8	537	4	US-09-925-065A-40292
C 37	20.2	80.8	601	5	US-10-027-632-303508
C 38	20.2	80.8	1089	4	US-09-925-065A-303508
C 39	20.2	80.8	1355	8	US-10-425-115-101300
C 40	20.2	80.8	1889	4	US-09-925-065A-13347
C 41	20.2	80.8	2031	8	US-10-425-115-79256
C 42	20.2	80.8	4254	10	US-11-097-143-4504
C 43	20.2	80.8	6051	8	US-10-723-860-3834
C 44	20.2	80.8	6746	10	US-11-097-143-2875
C 45	20.2	80.8	8333	5	US-10-239-676-113
C 46	20.2	80.8	8333	6	US-10-311-455-1475
C 47	20.2	80.8	8333	6	US-10-240-453-129
C 48	20.2	80.8	9483	6	US-10-311-455-350
C 49	20.2	80.8	9483	7	US-10-221-613-48
C 50	20.2	80.8	9483	7	US-10-087-192-1246
C 51	20.2	80.8	32433	7	US-10-052-482-136
C 52	20.2	80.8	42027	8	US-10-417-375-58
C 53	20.2	80.8	61007	7	US-10-221-714A-514
C 54	20.2	80.8	109730	8	US-10-741-600-17809
C 55	20.2	80.8	168976	6	US-10-085-117-139
C 56	20.2	80.8	235070	5	US-10-087-192-1246
C 57	20.2	80.8	659188	3	US-09-771-208-20
C 58	20.2	80.8	25	7	US-10-681-773-251
C 59	20	79.2	43	3	US-09-932-165-1485
C 60	19.8	79.2	43	3	US-09-942-052-714
C 61	19.8	79.2	43	3	US-10-121-019-13
C 62	19.8	79.2	43	6	US-10-013-312-2995
C 63	19.8	79.2	43	6	US-10-087-190-30
C 64	19.8	79.2	43	6	US-10-120-885A-28
C 65	19.8	79.2	43	6	US-10-121-016-55
C 66	19.8	79.2	43	6	US-10-114-669-6
C 67	19.8	79.2	43	7	US-10-114-632-42
C 68	19.8	79.2	43	7	US-10-114-632-42
C 69	19.8	79.2	124	4	US-09-771-312-7
C 70	19.8	79.2	144	5	US-10-121-019-4
C 71	19.8	79.2	192	9	US-10-837-869-1
C 72	19.8	79.2	317	8	US-10-425-115-67105
C 73	19.8	79.2	317	8	US-10-121-019-1
C 74	19.8	79.2	317	8	US-10-121-019-1
C 75	19.8	79.2	425	3	US-09-771-312-3
C 76	19.8	79.2	581	4	US-09-925-065A-937904
C 77	19.8	79.2	2000	4	US-09-925-065A-937905
C 78	19.8	79.2	2000	3	US-09-938-842A-2750
C 79	19.4	77.6	558	5	US-10-027-632-198354
C 80	19.4	77.6	558	5	US-09-943-676-102
C 81	19.2	76.8	50	3	US-09-766-553-102
C 82	19.2	76.8	50	3	US-09-438-486-102
C 83	19.2	76.8	50	3	US-10-053-758-102
C 84	19.2	76.8	50	5	US-10-054-295-102
C 85	19.2	76.8	50	5	US-10-054-611-102
C 86	19.2	76.8	50	6	US-10-325-810-587
C 87	19.2	76.8	50	8	US-10-877-122-587
C 88	19.2	76.8	50	8	US-10-877-122-587
C 89	19.2	76.8	50	8	US-10-877-122-587
C 90	19.2	76.8	52	3	US-10-877-122-587
C 91	19.2	76.8	52	3	US-09-745-605-9
C 92	19.2	76.8	52	5	US-10-154-517-1
C 93	19.2	76.8	52	5	US-10-154-517-2
C 94	19.2	76.8	52	5	US-10-158-735-17
C 95	19.2	76.8	52	8	US-10-468-488-473
C 96	19.2	76.8	52	8	US-10-475-962-47

Sequence 800109, A	Sequence 800110, A	Sequence 800111, A	Sequence 854017, A	Sequence 14574, A	Sequence 133939, A	Sequence 166, App	Sequence 1784, App	Sequence 188, App	Sequence 284, App	Sequence 4072, App	Sequence 40292, A	Sequence 303508, A	Sequence 303508, A	Sequence 554872, A	Sequence 101300, A	Sequence 13347, A	Sequence 79256, A	Sequence 4504, App	Sequence 3834, App	Sequence 2875, App	Sequence 113, App	Sequence 129, App	Sequence 350, App	Sequence 48, App	Sequence 1246, App	Sequence 136, App	Sequence 58, App	Sequence 1809, A	Sequence 139, App	Sequence 1990, App	Sequence 20, App	Sequence 251, App	Sequence 1485, App	Sequence 714, App	Sequence 11, App	Sequence 295, App	Sequence 30, App	Sequence 28, App	Sequence 55, App	Sequence 46, App	Sequence 4, App	Sequence 7, App	Sequence 1, App	Sequence 6105, A	Sequence 1, App	Sequence 937904, A	Sequence 937905, A	Sequence 2750, App	Sequence 198354, A	Sequence 102, App	Sequence 102, App	Sequence 102, App	Sequence 102, App	Sequence 102, App	Sequence 587, App	Sequence 587, App	Sequence 587, App	Sequence 587, App	Sequence 1, App	Sequence 9, App	Sequence 21, App	Sequence 17, App	Sequence 473, App	Sequence 47, App
--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	-------------------	--------------------	-------------------	-------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	-------------------	-------------------	--------------------	--------------------	--------------------	-------------------	-------------------	-------------------	------------------	--------------------	-------------------	------------------	------------------	-------------------	--------------------	------------------	-------------------	--------------------	-------------------	------------------	-------------------	------------------	------------------	------------------	------------------	-----------------	-----------------	-----------------	------------------	-----------------	--------------------	--------------------	--------------------	--------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-----------------	-----------------	------------------	------------------	-------------------	------------------

```

C 97 19.2 76.8 52 8 US-10-797-613-12 Sequence 12, Appl
C 98 19.2 76.8 52 9 US-10-916-782-17 Sequence 17, Appl
C 99 19.2 76.8 115 5 US-10-102-524-1482 Sequence 1482, Ap
C 100 19.2 76.8 135 8 US-10-357-930-62029 Sequence 62029, A
C 101 19.2 76.8 172 8 US-10-425-115-55214 Sequence 55214, A
C 102 19.2 76.8 201 8 US-10-719-993-39249 Sequence 39249, A
C 103 19.2 76.8 208 3 US-09-834-975-54 Sequence 54, Appl
C 104 19.2 76.8 307 8 US-10-425-115-14640 Sequence 14640,
C 105 19.2 76.8 308 8 US-10-425-115-142042 Sequence 142042,
C 106 19.2 76.8 331 8 US-10-425-115-94846 Sequence 94846, A
C 107 19.2 76.8 334 9 US-10-950-009-922 Sequence 922, App
C 108 19.2 76.8 344 8 US-10-425-115-85675 Sequence 85675, A
C 109 19.2 76.8 354 9 US-10-950-009-614 Sequence 614, App
C 110 19.2 76.8 370 4 US-09-925-065A-782593 Sequence 782593,
C 111 19.2 76.8 381 3 US-09-920-300A-480 Sequence 480, App
C 112 19.2 76.8 381 5 US-10-033-528-480 Sequence 480, App
C 113 19.2 76.8 381 6 US-10-099-926-480 Sequence 480, App
C 114 19.2 76.8 381 9 US-10-961-527-480 Sequence 480, App
C 115 19.2 76.8 387 8 US-10-357-930-7904 Sequence 7904, Ap
C 116 19.2 76.8 400 8 US-10-357-930-7442 Sequence 7442, Ap
C 117 19.2 76.8 412 6 US-10-125-968-31 Sequence 31, Appl
C 118 19.2 76.8 435 7 US-10-424-599-113640 Sequence 113640,
C 119 19.2 76.8 439 4 US-09-925-065A-639643 Sequence 639643,
C 120 19.2 76.8 450 7 US-10-424-599-18855 Sequence 18855, A
C 121 19.2 76.8 460 7 US-10-424-599-60015 Sequence 60015, A
C 122 19.2 76.8 521 8 US-10-357-930-57613 Sequence 57613, A
C 123 19.2 76.8 531 9 US-10-950-009-138 Sequence 138, App
C 124 19.2 76.8 532 4 US-09-925-065A-594044 Sequence 594044,
C 125 19.2 76.8 535 8 US-10-425-115-25718 Sequence 25718, A
C 126 19.2 76.8 548 4 US-09-925-065A-79960 Sequence 79960, A
C 127 19.2 76.8 553 7 US-10-663-561-57 Sequence 57, Appl
C 128 19.2 76.8 555 4 US-09-925-065A-43740 Sequence 43740, A
C 129 19.2 76.8 555 4 US-09-925-065A-43741 Sequence 43741, A
C 130 19.2 76.8 558 4 US-09-925-065A-43742 Sequence 43742, A
C 131 19.2 76.8 558 8 US-10-425-115-64082 Sequence 64082, A
C 132 19.2 76.8 562 4 US-09-925-065A-559065 Sequence 559065,
C 133 19.2 76.8 562 4 US-09-925-065A-559066 Sequence 559066,
C 134 19.2 76.8 562 4 US-09-925-065A-559067 Sequence 559067,
C 135 19.2 76.8 566 4 US-09-925-065A-56190 Sequence 56190, A
C 136 19.2 76.8 567 4 US-09-925-065A-600458 Sequence 600458,
C 137 19.2 76.8 567 4 US-09-925-065A-600459 Sequence 600459,
C 138 19.2 76.8 567 4 US-09-925-065A-600460 Sequence 600460,
C 139 19.2 76.8 567 4 US-09-925-065A-600461 Sequence 600461,
C 140 19.2 76.8 569 4 US-09-925-065A-254763 Sequence 254763,
C 141 19.2 76.8 566 4 US-09-925-065A-272531 Sequence 272531,
C 142 19.2 76.8 591 4 US-09-925-065A-272531 Sequence 272531,
C 143 19.2 76.8 594 8 US-10-425-115-32954 Sequence 32954, A
C 144 19.2 76.8 606 4 US-09-925-065A-737130 Sequence 737130,
C 145 19.2 76.8 606 4 US-09-925-065A-743446 Sequence 743446,
C 146 19.2 76.8 606 4 US-09-925-065A-814497 Sequence 814497,
C 147 19.2 76.8 611 5 US-10-027-632-194033 Sequence 194033,
C 148 19.2 76.8 611 6 US-10-027-632-194033 Sequence 194033,
C 149 19.2 76.8 620 8 US-10-425-115-130411 Sequence 130411,
C 150 19.2 76.8 624 4 US-09-925-065A-469768 Sequence 469768,

```

ALIGNMENTS

```

RESULT 1
US-10-681-773-3
; Sequence 3, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475

```

```

; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-3

```

```

Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 1 AAAAAAAAACTAAAGCTTGATCTT 25

```

```

RESULT 2
US-10-027-632-63078
; Sequence 63078, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63078
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63078

```

```

Query Match 98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 370 AAAAAAAAACTAAAGCTTGATCTT 394

```

```

RESULT 3
US-10-027-632-63079
; Sequence 63079, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30

```

```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63079
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63079
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTAAAGCTTGATCTT 25
         |||||:|||||:|||||:|||||
Db       370 AAAAAAAAACTAAGCTTGATCTT 394
```

```

RESULT 4
US-10-027-632-63080
; Sequence 63080, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63080
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63080
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTAAAGCTTGATCTT 25
         |||||:|||||:|||||:|||||
Db       370 AAAAAAAAACTAAGCTTGATCTT 394
```

```

RESULT 5
US-10-027-632-63081
; Sequence 63081, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63081
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63081
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTAAAGCTTGATCTT 25
         |||||:|||||:|||||:|||||
Db       370 AAAAAAAAACTAAGCTTGATCTT 394
```

```

RESULT 6
US-10-027-632-63078
; Sequence 63078, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63078
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63078
```

US-10-027-632-63078

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

RESULT 7

US-10-027-632-63079
; Sequence 63079, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63079.
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63079

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

RESULT 8

US-10-027-632-63080
; Sequence 63080, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63080
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63080

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 9

US-10-027-632-63081
; Sequence 63081, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63081
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63081

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

QY 1 AAAAAAAAACTAAGCTGATCTT 25
|||||:|||||:
DB 370 AAAAAAAAACTAAGCTGATCTT 394

```

; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 10
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-10

Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 1 AAAAAAAAACTAAGCTTGATCTT 24

RESULT 11
US-10-681-773-87004
; Sequence 87004, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 87004
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-87004

Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 2 AAAAAAAAACTAAGCTTGATCTT 25

RESULT 12
US-10-027-632-39077
; Sequence 39077, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```

; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39077
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39077

Query Match          94.4%; Score 23.6; DB 5; Length 501;
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAGCTTGATCTT 370

RESULT 13
US-10-027-632-39078
; Sequence 39078, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39078
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39078

Query Match          94.4%; Score 23.6; DB 5; Length 501;
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAGCTTGATCTT 370

RESULT 14
US-10-027-632-39079
; Sequence 39079, Application US/10027632
```

```
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1999-09-28
PRIORITY APPLICATION NUMBER: US 60/146,002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39079
LENGTH: 501
TYPE: DNA
ORGANISM: Human
US-10-027-632-39079
```

```
Query Match          94.4%; Score 23.6; DB 5; Length 501;
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 2 AAAAAAAAACTAAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAAGCTTGATCTT 370
```

```
RESULT 15
US-10-027-632-39080
Sequence 39080, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1999-09-28
PRIORITY APPLICATION NUMBER: US 60/146,002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39080
LENGTH: 501
TYPE: DNA
ORGANISM: Human
US-10-027-632-39080
```

```
Query Match          94.4%; Score 23.6; DB 5; Length 501;
```

```
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 2 AAAAAAAAACTAAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAAGCTTGATCTT 370
```

```
RESULT 16
US-10-027-632-39077
Sequence 39077, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1999-09-28
PRIORITY APPLICATION NUMBER: US 60/146,002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39077
LENGTH: 501
TYPE: DNA
ORGANISM: Human
US-10-027-632-39077
```

```
Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 2 AAAAAAAAACTAAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAAGCTTGATCTT 370
```

```
RESULT 17
US-10-027-632-39078
Sequence 39078, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1999-09-28
```

```

; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39078
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39078

Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAGCTTGATCTT 370

RESULT 18
US-10-027-632-39079
; Sequence 39079, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39079
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39079

Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAGCTTGATCTT 370

RESULT 19
US-10-027-632-39080
; Sequence 39080, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
```

```

; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39080
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39080

Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 32;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 347 AAAAAAAAACTAAGCTTGATCTT 370

RESULT 20
US-10-681-773-4
; Sequence 4, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 4
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-4

Query Match          93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 23;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 1 AAAAAAAAACTAAGCTTGATCTT 25

RESULT 21
US-10-681-773-20
; Sequence 20, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
```

```

; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 20
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-20

Query Match          92.0%; Score 23; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAAAAAAAACTAAGCTTGATCTT 25
Db 1 AAAAAAAAACTAAGCTTGATCTT 23

RESULT 22
US-10-681-773-11
; Sequence 11, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 11
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-11

Query Match          89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 56;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 1 AAAAAAAAACTAAGCTTGATCTT 24

RESULT 23
US-10-681-773-87005
; Sequence 87005, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```

; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 87005
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-87005

Query Match          89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 56;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 2 AAAAAAAAACTAAGCTTGATCT 25

RESULT 24
US-09-925-065A-800109/c
; Sequence 800109, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800109
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-800109

Query Match          89.6%; Score 22.4; DB 4; Length 628;
Best Local Similarity 95.8%; Pred. No. 96;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
Db 281 AAAAAAAAACTAAGCTTGATCT 258

RESULT 25
US-09-925-065A-800110/c
; Sequence 800110, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
```



```

; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800110
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-800110

Query Match      89.6%; Score 22.4; DB 4; Length 628;
Best Local Similarity 95.8%; Pred. No. 96;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 24
Db 281 AAAAAAAAACTAAGCTTGATCT 258

RESULT 26
; Sequence 800111, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800111
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-800111

Query Match      89.6%; Score 22.4; DB 4; Length 628;
Best Local Similarity 95.8%; Pred. No. 96;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 24
Db 281 AAAAAAAAACTAAGCTTGATCT 258

RESULT 27
; Sequence 854017, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
```

```

; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 854017
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-854017

Query Match      89.6%; Score 22.4; DB 4; Length 628;
Best Local Similarity 95.8%; Pred. No. 96;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 24
Db 281 AAAAAAAAACTAAGCTTGATCT 258

RESULT 28
; Sequence 21, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 21
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
; US-10-681-773-21

Query Match      85.6%; Score 21.4; DB 7; Length 25;
Best Local Similarity 95.7%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAAAAAAAACTAAGCTTGATCT 25
Db 1 AAAAAAAAACTAAGCTTGATCT 23

RESULT 29
; Sequence 14574, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

```

; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 14574
; LENGTH: 1666
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB23-006-F9_FLI
US-10-425-114-14574

Query Match      85.6%; Score 21.4; DB 7; Length 1666;
Best Local Similarity 95.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3  AAAAAAAAACTAAGCTTGATCTT 25
DB      1643  AAAAAAGACTAAGCTTGATCTT 1665

RESULT 30
US-09-925-065A-334939/c
; Sequence 334939, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334939
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-334939

Query Match      83.2%; Score 20.8; DB 4; Length 521;
Best Local Similarity 91.7%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  AAAAAAAAACTAAGCTTGATCT 24
DB      38  AAAAAAAAACTAAGCTTGATCT 15

RESULT 31
US-10-239-676-166/c
; Sequence 166, Application US/10239676
; Publication No. US2003082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8

; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 166
; LENGTH: 13606
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (5950, 5973, 8958)
US-10-239-676-166

Query Match      83.2%; Score 20.8; DB 5; Length 13606;
Best Local Similarity 91.7%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  AAAAAAAAACTAAGCTTGATCT 24
DB      5439  AAAAAAAAACTAAGCTTGATCT 5416

RESULT 32
US-10-311-455-1784/c
; Sequence 1784, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1784
; LENGTH: 13606
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: 5950, 5973, 8958
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1784

Query Match      83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  AAAAAAAAACTAAGCTTGATCT 24
DB      5439  AAAAAAAAACTAAGCTTGATCT 5416

RESULT 33
US-10-240-453-188/c
; Sequence 188, Application US/10240453
```

```
Publication No. US20030148326A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
TITLE OF INVENTION: Transcription
TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
FILE REFERENCE: 5013.1009
CURRENT APPLICATION NUMBER: US/10/240.453
CURRENT FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: PCT/EP01/03973
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 350
SEQ ID NO 188
LENGTH: 13606
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (5950, 5973, 8958)
US-10-240-453-188

Query Match      83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCT 24
Db      5439 AAAAAAAAACTAAGCTTATCT 5416

RESULT 34
US-10-221-714A-284/C
Sequence 284, Application US/10221714A
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with
FILE REFERENCE: 5013.1005
CURRENT APPLICATION NUMBER: US/10/221.714A
CURRENT FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: PCT/EP01/02955
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: DE 10013847.0
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 540
SEQ ID NO 284
LENGTH: 13606
TYPE: DNA
ORGANISM: Artificial Sequence
```

```
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: 5950, 5973, 8958
OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-284

Query Match      83.2%; Score 20.8; DB 7; Length 13606;
Best Local Similarity 91.7%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCT 24
Db      5439 AAAAAAAAACTAAGCTTATCT 5416

RESULT 35
US-10-357-930-4072/C
Sequence 4072, Application US/10357930
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Endege, Wilson
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
FILE REFERENCE: MRI-007BCN
CURRENT APPLICATION NUMBER: US/10/357.930
CURRENT FILING DATE: 2003-02-04
PRIOR APPLICATION NUMBER: 09/785,276
PRIOR FILING DATE: 2003-02-16
PRIOR APPLICATION NUMBER: 60/183,319
PRIOR FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/189,862
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/207,454
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: 60/211,314
PRIOR FILING DATE: 2000-06-09
PRIOR APPLICATION NUMBER: 60/219,007
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/255,281
PRIOR FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 62232
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4072
LENGTH: 424
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 6, 39, 385
OTHER INFORMATION: n = A,T,C or G
US-10-357-930-4072

Query Match      80.8%; Score 20.2; DB 8; Length 424;
Best Local Similarity 88.0%; Pred. No. 6.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCT 25
Db      374 AAAAAAAAACTAAGCTTATCTT 350

RESULT 36
US-09-925-065A-40292/C
Sequence 40292, Application US/09925065A
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
```

```

; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40292
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-40292

Query Match      80.8%; Score 20.2; DB 4; Length 537;
Best Local Similarity 88.0%; Pred. No. 6.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTGATCTT 25
Db 142 AAAAAAAAAAAAGCTGACCTT 118

RESULT 37
US-10-027-632-303508
; Sequence 303508, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 303508
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(601)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-303508

Query Match      80.8%; Score 20.2; DB 5; Length 601;
Best Local Similarity 88.0%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTGATCTT 25
Db 1 AAAAAAAAACTAAAGCTGATCTT 25
```

```

Db 48 AAAAAAAAAATTAAAGCTGATCTT 72

RESULT 38
US-10-027-632-303508
; Sequence 303508, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 303508
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(601)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-303508

Query Match      80.8%; Score 20.2; DB 6; Length 601;
Best Local Similarity 88.0%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTGATCTT 25
Db 48 AAAAAAAAAATTAAAGCTGATCTT 72

RESULT 39
US-09-925-065A-554872
; Sequence 554872, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 554872
; LENGTH: 1089
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-554872

Query Match      80.8%; Score 20.2; DB 4; Length 1089;
Best Local Similarity 88.0%; Pred. No. 7.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 569 AAAAAAAAACTAAGCTTGATCTT 593

RESULT 40
US-10-425-115-101300
; Sequence 101300, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 101300
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_2389C.1
US-10-425-115-101300

Query Match      80.8%; Score 20.2; DB 8; Length 1355;
Best Local Similarity 88.0%; Pred. No. 7.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 34 AAAAAAAAACTAAGCTTGATCTT 58

RESULT 41
US-09-925-065A-13347
; Sequence 13347, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13347
; LENGTH: 1889
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
US-09-925-065A-13347

Query Match      80.8%; Score 20.2; DB 4; Length 1889;
Best Local Similarity 88.0%; Pred. No. 8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 1071 AAAAAAAAACTAAGCTTGATCTT 1095

RESULT 42
US-10-425-115-79256
; Sequence 79256, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 79256
; LENGTH: 2031
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_172307C.1
US-10-425-115-79256
```

```

Query Match      80.8%; Score 20.2; DB 8; Length 2031;
Best Local Similarity 88.0%; Pred. No. 8.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 1789 AAAAAAAAACTAAGCTTGATCTT 1813

RESULT 43
US-11-097-143-4504/C
; Sequence 4504, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
```

```

; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4504
; LENGTH: 4254
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-4504

Query Match      80.8%; Score 20.2; DB 10; Length 4254;
Best Local Similarity 88.0%; Pred. No. 9.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 827 AAAAAACAAACTAAGCTCGATATT 803

RESULT 44
US-10-723-860-3834/C
; Sequence 3834, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3834
; LENGTH: 6051
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-3834

Query Match      80.8%; Score 20.2; DB 8; Length 6051;
Best Local Similarity 88.0%; Pred. No. 9.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 4168 AAAAAACAAACAAAGGATGATCTT 4144

RESULT 45
US-11-097-143-2875/C
; Sequence 2875, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693

; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2875
; LENGTH: 6746
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-2875

Query Match      80.8%; Score 20.2; DB 10; Length 6746;
Best Local Similarity 88.0%; Pred. No. 9.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 5953 AAAAAACAAACTAAGCTCGATATT 5929

RESULT 46
US-10-239-676-113/C
; Sequence 113, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 113
; LENGTH: 8333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-113

Query Match      80.8%; Score 20.2; DB 5; Length 8333;
Best Local Similarity 88.0%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 525 AAAAAACAAACTAAGCTTAACCTT 501

RESULT 47
US-10-311-455-1475/C
; Sequence 1475, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
```

```

; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1475
; LENGTH: 8333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1475

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 6; Length 8333;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 525 AAAAAAAAACTAAGCTTGATCTT 501

RESULT 48
US-10-240-453-129/c
; Sequence 129, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: with DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 129
; LENGTH: 8333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-129

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 6; Length 8333;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 525 AAAAAAAAACTAAGCTTGATCTT 501

RESULT 49
US-10-311-455-350/c
; Sequence 350, Application US/10311455
; Publication No. US20030143606A1

; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 350
; LENGTH: 9483
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-350

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 6; Length 9483;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 4182 AAAAAAAAACTAAGCTTGATCTT 4158

RESULT 50
US-10-221-613-48/c
; Sequence 48, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221.613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; PRIOR FILING DATE: 2001-03-15
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 48
; LENGTH: 9483
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-48

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 7; Length 9483;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 4182 AAAAAAAAACTAAGCTTGATCTT 4158
```

Thu Dec 15 09:19:11 2005

us-10-681-773-3.rnpbm

Page 16

Search completed: December 14, 2005, 08:46:18
Job time : 376.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 : Search time 373.2 Seconds

(without alignments)
553.951 Million cell updates/sec

Title: US-10-681-773-4

Perfect score: 25
Sequence: 1 aaaaaaaaaactagctgcttc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 413468905 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	7	US-10-681-773-4
2	24.6	98.4	535	5	US-10-027-632-63078
3	24.6	98.4	535	5	US-10-027-632-63079
4	24.6	98.4	535	5	US-10-027-632-63080
5	24.6	98.4	535	5	US-10-027-632-63081
6	24.6	98.4	535	5	US-10-027-632-63078
7	24.6	98.4	535	5	US-10-027-632-63079
8	24.6	98.4	535	5	US-10-027-632-63080
9	24.6	98.4	535	5	US-10-027-632-63081
10	24	96.0	25	7	US-10-681-773-11
11	24	96.0	25	7	US-10-681-773-87005
12	23.6	94.4	501	5	US-10-027-632-39077
13	23.6	94.4	501	5	US-10-027-632-39078
14	23.6	94.4	501	5	US-10-027-632-39079
15	23.6	94.4	501	5	US-10-027-632-39080
16	23.6	94.4	501	5	US-10-027-632-39077
17	23.6	94.4	501	5	US-10-027-632-39078
18	23.6	94.4	501	5	US-10-027-632-39079
19	23.6	94.4	501	5	US-10-027-632-39080
20	23.6	94.4	501	5	US-10-027-632-39077
21	23.6	94.4	501	5	US-10-027-632-39078
22	23.6	94.4	501	5	US-10-027-632-39079
23	23.6	94.4	501	5	US-10-027-632-39080

24	21.4	85.6	25	7	US-10-681-773-20	Sequence 20, Appl
25	20.8	83.2	628	4	US-09-925-065A-800109	Sequence 800109,
26	20.8	83.2	628	4	US-09-925-065A-800110	Sequence 800110,
27	20.8	83.2	628	4	US-09-925-065A-800111	Sequence 800111,
28	20.8	83.2	628	4	US-09-925-065A-854017	Sequence 854017,
29	20.4	81.6	595	4	US-09-925-065A-131554	Sequence 131554,
30	20.2	80.8	210	8	US-10-425-115-179221	Sequence 179221,
31	20.2	80.8	5228	6	US-10-311-455-1627	Sequence 1627, Ap
32	20	80.0	25	7	US-10-681-773-258	Sequence 258, App
33	19.8	79.2	1666	7	US-10-425-114-14574	Sequence 14574, A
34	19.4	77.6	2678276	5	US-10-087-192-754	Sequence 86064, A
35	19.4	77.6	2678276	5	US-10-087-192-754	Sequence 754, App
36	19.2	76.8	335	7	US-10-437-963-42936	Sequence 42936, A
37	19.2	76.8	405	8	US-10-425-115-144586	Sequence 144586,
38	19.2	76.8	469	9	US-10-915-740A-107	Sequence 107, App
39	19.2	76.8	469	9	US-10-915-740A-107	Sequence 65854, A
40	19.2	76.8	521	4	US-09-925-065A-134939	Sequence 314939,
41	19.2	76.8	521	4	US-09-925-065A-134939	Sequence 314939, A
42	19.2	76.8	522	4	US-09-925-065A-59437	Sequence 59437,
43	19.2	76.8	618	4	US-09-925-065A-264897	Sequence 264897,
44	19.2	76.8	669	4	US-09-925-065A-870597	Sequence 870597,
45	19.2	76.8	788	5	US-10-027-632-146140	Sequence 146140,
46	19.2	76.8	788	5	US-10-027-632-146140	Sequence 146140,
47	19.2	76.8	2123	6	US-10-369-493-29566	Sequence 29566, A
48	19.2	76.8	2733	4	US-09-925-065A-711163	Sequence 711163,
49	19.2	76.8	6141	10	US-11-097-143-10747	Sequence 10747, A
50	19.2	76.8	6247	7	US-10-257-166-4	Sequence 4, Appl
51	19.2	76.8	7495	10	US-11-097-143-19351	Sequence 19351, A
52	19.2	76.8	13606	5	US-10-339-676-166	Sequence 166, App
53	19.2	76.8	13606	5	US-10-311-455-1784	Sequence 1784, App
54	19.2	76.8	13606	6	US-10-240-453-188	Sequence 188, App
55	19.2	76.8	13606	6	US-10-221-714A-284	Sequence 284, App
56	19.2	76.8	14032	6	US-10-311-455-1425	Sequence 1425, Ap
57	19.2	76.8	23349	8	US-10-719-993-7038	Sequence 6841, Ap
58	19.2	76.8	54738	8	US-10-719-993-7038	Sequence 7038, Ap
59	19.2	76.8	191597	8	US-10-424-963-6802	Sequence 6802, Ap
60	18.8	75.2	226	7	US-10-424-963-6802	Sequence 15967, A
61	18.8	75.2	460	3	US-09-764-855-230	Sequence 230, App
62	18.8	75.2	460	3	US-10-072-349-230	Sequence 230, App
63	18.8	75.2	460	3	US-09-764-855-58	Sequence 58, Appl
64	18.8	75.2	469	5	US-10-072-349-58	Sequence 58, Appl
65	18.8	75.2	537	7	US-10-021-323-15954	Sequence 15954, A
66	18.8	75.2	581	4	US-09-925-065A-127992	Sequence 127992,
67	18.8	75.2	581	4	US-09-925-065A-127993	Sequence 127993,
68	18.8	75.2	587	4	US-09-925-065A-310495	Sequence 310495,
69	18.8	75.2	591	4	US-09-925-065A-611374	Sequence 611374,
70	18.8	75.2	1312	5	US-10-027-632-123984	Sequence 123984,
71	18.8	75.2	1312	5	US-10-027-632-123985	Sequence 123985,
72	18.8	75.2	1312	6	US-10-027-632-123984	Sequence 123984,
73	18.8	75.2	1312	6	US-10-027-632-123985	Sequence 123985,
74	18.8	75.2	1805	4	US-09-925-065A-695530	Sequence 695530,
75	18.8	75.2	1805	4	US-09-925-065A-695531	Sequence 695531,
76	18.8	75.2	1805	4	US-09-925-065A-695532	Sequence 695532,
77	18.8	75.2	1805	4	US-09-925-065A-695533	Sequence 695533,
78	18.8	75.2	1805	4	US-09-925-065A-695534	Sequence 695534,
79	18.8	75.2	1805	4	US-09-925-065A-695535	Sequence 695535,
80	18.8	75.2	1805	4	US-09-925-065A-695536	Sequence 695536,
81	18.8	75.2	1805	4	US-09-925-065A-695537	Sequence 695537,
82	18.8	75.2	1865	7	US-10-437-963-50036	Sequence 73270, A
83	18.8	75.2	2081	7	US-10-437-963-50036	Sequence 50036, A
84	18.8	75.2	7448	7	US-10-421-714A-247	Sequence 247, App
85	18.8	75.2	7448	7	US-10-421-714A-247	Sequence 58, Appl
86	18.8	75.2	7703	6	US-10-311-455-1408	Sequence 1408, Ap
87	18.8	75.2	9741	6	US-10-311-455-1408	Sequence 1406, Ap
88	18.8	75.2	152744	9	US-10-981-277-41	Sequence 41, Appl
89	18.8	75.2	209083	9	US-10-461-86A-74	Sequence 74, Appl
90	18.6	74.4	235	7	US-10-242-532A-53842	Sequence 53842, A
91	18.6	74.4	235	7	US-10-085-781A-53842	Sequence 53842, A
92	18.6	74.4	251	8	US-10-425-115-184542	Sequence 184542,
93	18.6	74.4	263	7	US-10-424-599-133907	Sequence 133907,
94	18.6	74.4	285	8	US-10-425-115-138330	Sequence 38830, A
95	18.6	74.4	312	8	US-10-357-930-6598	Sequence 6598, Ap
96	18.6	74.4	341	7	US-10-424-599-77833	Sequence 77833, A

```
C 97 18.6 74.4 391 8 US-10-357-930-36574 Sequence 36574, A
C 98 18.6 74.4 424 8 US-10-357-930-4072 Sequence 4072, A
C 99 18.6 74.4 445 5 US-09-925-065A-230605 Sequence 230605, A
C 100 18.6 74.4 446 5 US-10-027-632-43399 Sequence 43399, A
C 101 18.6 74.4 446 6 US-10-027-632-43399 Sequence 43399, A
C 102 18.6 74.4 445 4 US-09-925-065A-44467 Sequence 44467, A
C 103 18.6 74.4 455 4 US-09-925-065A-453497 Sequence 453497, A
C 104 18.6 74.4 472 4 US-09-925-065A-801284 Sequence 801284, A
C 105 18.6 74.4 477 4 US-09-925-065A-780291 Sequence 780291, A
C 106 18.6 74.4 480 7 US-09-925-065A-418376 Sequence 418376, A
C 107 18.6 74.4 480 7 US-10-424-599-16604 Sequence 16604, A
C 108 18.6 74.4 482 5 US-10-027-632-72395 Sequence 72395, A
C 109 18.6 74.4 482 6 US-10-027-632-72395 Sequence 72395, A
C 110 18.6 74.4 485 5 US-10-027-632-312707 Sequence 312707, A
C 111 18.6 74.4 485 6 US-10-027-632-312707 Sequence 312707, A
C 112 18.6 74.4 503 4 US-09-925-065A-769670 Sequence 769670, A
C 113 18.6 74.4 513 7 US-10-424-599-930 Sequence 930, A
C 114 18.6 74.4 526 4 US-09-925-065A-17529 Sequence 17529, A
C 115 18.6 74.4 526 4 US-09-925-065A-17529 Sequence 17529, A
C 116 18.6 74.4 526 4 US-09-925-065A-17530 Sequence 17530, A
C 117 18.6 74.4 537 4 US-09-925-065A-40292 Sequence 40292, A
C 118 18.6 74.4 539 4 US-09-925-065A-135534 Sequence 135534, A
C 119 18.6 74.4 550 5 US-10-027-632-531 Sequence 531, A
C 120 18.6 74.4 550 5 US-10-027-632-532 Sequence 532, A
C 121 18.6 74.4 550 6 US-10-027-632-531 Sequence 531, A
C 122 18.6 74.4 550 6 US-10-027-632-532 Sequence 532, A
C 123 18.6 74.4 553 4 US-09-925-065A-184968 Sequence 184968, A
C 124 18.6 74.4 572 5 US-10-027-632-224078 Sequence 224078, A
C 125 18.6 74.4 572 6 US-10-027-632-224078 Sequence 224078, A
C 126 18.6 74.4 576 4 US-09-925-065A-196716 Sequence 196716, A
C 127 18.6 74.4 577 4 US-09-925-065A-184969 Sequence 184969, A
C 128 18.6 74.4 581 5 US-10-027-632-309042 Sequence 309042, A
C 129 18.6 74.4 591 6 US-10-027-632-309042 Sequence 309042, A
C 130 18.6 74.4 594 4 US-09-925-065A-160406 Sequence 160406, A
C 131 18.6 74.4 596 4 US-09-925-065A-257438 Sequence 257438, A
C 132 18.6 74.4 600 9 US-10-972-079-90545 Sequence 90545, A
C 133 18.6 74.4 601 6 US-10-027-632-303508 Sequence 303508, A
C 134 18.6 74.4 601 6 US-10-027-632-303508 Sequence 303508, A
C 135 18.6 74.4 603 4 US-09-925-065A-778257 Sequence 778257, A
C 136 18.6 74.4 606 4 US-09-925-065A-746051 Sequence 746051, A
C 137 18.6 74.4 610 4 US-09-925-065A-67879 Sequence 67879, A
C 138 18.6 74.4 610 4 US-09-925-065A-357338 Sequence 357338, A
C 139 18.6 74.4 615 4 US-09-925-065A-672208 Sequence 672208, A
C 140 18.6 74.4 623 5 US-10-027-632-276707 Sequence 276707, A
C 141 18.6 74.4 623 6 US-10-027-632-276707 Sequence 276707, A
C 142 18.6 74.4 633 4 US-09-925-065A-445241 Sequence 445241, A
C 143 18.6 74.4 633 4 US-09-925-065A-445242 Sequence 445242, A
C 144 18.6 74.4 633 4 US-09-925-065A-445243 Sequence 445243, A
C 145 18.6 74.4 636 5 US-10-027-632-43268 Sequence 43268, A
C 146 18.6 74.4 636 6 US-10-027-632-43268 Sequence 43268, A
C 147 18.6 74.4 664 7 US-10-425-114-28139 Sequence 28139, A
C 148 18.6 74.4 703 8 US-10-363-345A-22435 Sequence 22435, A
C 149 18.6 74.4 703 8 US-10-363-345A-22436 Sequence 22436, A
C 150 18.6 74.4 703 9 US-10-363-345A-22435 Sequence 22435, A
```

ALIGNMENTS

```
RESULT 1
US-10-681-773-4
; Sequence 4, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 4
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-4
```

```
Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 1 AAAAAAAAACTATAGCTTGATCTT 25
```

```
RESULT 2
US-10-027-632-63078
; Sequence 63078, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63078
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63078
```

```
Query Match 98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 370 AAAAAAAAACTATAGCTTGATCTT 394
```

```
RESULT 3
US-10-027-632-63079
; Sequence 63079, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```

```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63079
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63079
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 370 AAAAAAAAACTATAGCTTGATCTT 394
```

```

RESULT 4
US-10-027-632-63080
; Sequence 63080, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63080
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63080
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 370 AAAAAAAAACTATAGCTTGATCTT 394
```

```

RESULT 5
US-10-027-632-63081
; Sequence 63081, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63081
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63081
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 370 AAAAAAAAACTATAGCTTGATCTT 394
```

```

RESULT 6
US-10-027-632-63078
; Sequence 63078, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63078
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-63078

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 370 AAAAAAAAACTAGCTTGATCTT 394

RESULT 7

US-10-027-632-63079
; Sequence 63079, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63079
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63079

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 370 AAAAAAAAACTAGCTTGATCTT 394

RESULT 8

US-10-027-632-63080
; Sequence 63080, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 63080

; LENGTH: 525

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-63080

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 370 AAAAAAAAACTAGCTTGATCTT 394

RESULT 9

US-10-027-632-63081
; Sequence 63081, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63081
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63081

Query Match 98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 10;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 370 AAAAAAAAACTAGCTTGATCTT 394

RESULT 10

US-10-681-773-11
; Sequence 11, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia

```
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 11
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-11

Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATGCTTGATCTT 25
Db 1 AAAAAAAAACTATGCTTGATCTT 24

RESULT 11
US-10-681-773-87005
; Sequence 87005, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 87005
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-87005

Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATGCTTGATCTT 24
Db 2 AAAAAAAAACTATGCTTGATCTT 25

RESULT 12
US-10-027-632-39077
; Sequence 39077, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39077
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39077

Query Match          94.4%; Score 23.6; DB 5; Length 501;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATGCTTGATCTT 25
Db 347 AAAAAAAAACTATGCTTGATCTT 370

RESULT 13
US-10-027-632-39078
; Sequence 39078, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39078
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39078

Query Match          94.4%; Score 23.6; DB 5; Length 501;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATGCTTGATCTT 25
Db 347 AAAAAAAAACTATGCTTGATCTT 370

RESULT 14
US-10-027-632-39079
; Sequence 39079, Application US/10027632
```

```
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39079
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39079
```

```
Query Match          94.4%; Score 23.6; DB 5; Length 501;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 AAAAAAAAACTATAGCTTGATCTT 25
Db      347 AAAAAAAAACTATAGCTTGATCTT 370
```

```
RESULT 15
US-10-027-632-39080
; Sequence 39080, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39080
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39080
```

```
Query Match          94.4%; Score 23.6; DB 5; Length 501;
```

```
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 AAAAAAAAACTATAGCTTGATCTT 25
Db      347 AAAAAAAAACTATAGCTTGATCTT 370
```

```
RESULT 16
US-10-027-632-39077
; Sequence 39077, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39077
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39077
```

```
Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 AAAAAAAAACTATAGCTTGATCTT 25
Db      347 AAAAAAAAACTATAGCTTGATCTT 370
```

```
RESULT 17
US-10-027-632-39078
; Sequence 39078, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
```

```

; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39078
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39078

Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 347 AAAAAAAAACTAWAGCTTGATCTT 370

RESULT 18
US-10-027-632-39079
; Sequence 39079, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39079
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39079

Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 347 AAAAAAAAACTAWAGCTTGATCTT 370

RESULT 19
US-10-027-632-39080
; Sequence 39080, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
```

```

; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39080
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39080

Query Match          94.4%; Score 23.6; DB 6; Length 501;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 347 AAAAAAAAACTAWAGCTTGATCTT 370

RESULT 20
US-10-681-773-3
; Sequence 3, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-3

Query Match          93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 18;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 1 AAAAAAAAACTAWAGCTTGATCTT 25

RESULT 21
US-10-681-773-21
; Sequence 21, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
```

```
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ PRIOR FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO 21
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-21

Query Match          92.0%; Score 23; DB 7; Length 25;
Best Local Similarity 100.0%; Pred.No.26;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 1 AAAAAAAAACTATAGCTTGATCTT 23

RESULT 22
US-10-681-773-10
/ Sequence 10, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsuzaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO 10
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-10

Query Match          89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred.No.45;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 1 AAAAAAAAACTAAGCTTGATCTT 24

RESULT 23
US-10-681-773-87004
/ Sequence 87004, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsuzaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
```

```
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO 87004
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-87004

Query Match          89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred.No.45;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
    |||||
Db 2 AAAAAAAAACTAAGCTTGATCT 25

RESULT 24
US-10-681-773-20
/ Sequence 20, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsuzaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO 20
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-20

Query Match          85.6%; Score 21.4; DB 7; Length 25;
Best Local Similarity 95.7%; Pred.No.11e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 1 AAAAAAAAACTAAGCTTGATCTT 23

RESULT 25
US-09-925-065A-800109/c
/ Sequence 800109, Application US/09925065A
/ Publication No. US20050228172A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
/ FILE REFERENCE: 108827.135
/ CURRENT APPLICATION NUMBER: US/09/925,065A
/ CURRENT FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243,096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
```



```
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800109
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-800109
```

```
Query Match      83.2%; Score 20.8; DB 4; Length 628;
Best Local Similarity 91.7%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTAAGCTTGATCT 24
          |||||
Db      281 AAAAAAAAAAGCTTAAGCTTGATCT 258
```

RESULT 26

```
US-09-925-065A-800110/c
; Sequence 800110, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800110
; LENGTH: 628.
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-800110
```

```
Query Match      83.2%; Score 20.8; DB 4; Length 628;
Best Local Similarity 91.7%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTAAGCTTGATCT 24
          |||||
Db      281 AAAAAAAAAAGCTTAAGCTTGATCT 258
```

```
RESULT 27
US-09-925-065A-800111/c
; Sequence 800111, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
```

```
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800111
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-800111
```

```
Query Match      83.2%; Score 20.8; DB 4; Length 628;
Best Local Similarity 91.7%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTAAGCTTGATCT 24
          |||||
Db      281 AAAAAAAAAAGCTTAAGCTTGATCT 258
```

RESULT 28

```
US-09-925-065A-854017/c
; Sequence 854017, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 854017
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-854017
```

```
Query Match      83.2%; Score 20.8; DB 4; Length 628;
Best Local Similarity 91.7%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTAAGCTTGATCT 24
          |||||
Db      281 AAAAAAAAAAGCTTAAGCTTGATCT 258
```

```
RESULT 29
US-09-925-065A-131554
; Sequence 131554, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
```

```
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289,846
/ PRIOR FILING DATE: 2001-05-09
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 131554
/ LENGTH: 595
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-925-065A-131554

Query Match
Best Local Similarity 81.6%; Score 20.4; DB 4; Length 595;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGAT 22
    |||||
Db 492 AAAAAAAAACTATAGCTTTAT 513

RESULT 30
US-10-425-115-179221/C
/ Sequence 179221, Application US/10425115
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 179221
/ LENGTH: 210
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_95031C.1
US-10-425-115-179221

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 8; Length 210;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 46 AAAAAAAAAATATAGCTTGTTCTT 22

RESULT 31
US-10-311-455-1627/C
/ Sequence 1627, Application US/10311455
/ Publication No. US20030143606A1
/ GENERAL INFORMATION:
/ APPLICANT: OLEK, Alexander
/ APPLICANT: PIEPENBROCK, Christlan
/ APPLICANT: BERLIN, Kurt
/ TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
/ TITLE OF INVENTION: cytosine methylation
/ FILE REFERENCE: 5013.1014
/ CURRENT APPLICATION NUMBER: US/10/311,455
/ CURRENT FILING DATE: 2002-12-16
/ PRIOR APPLICATION NUMBER: PCT/EP01/07537
/ PRIOR FILING DATE: 2001-07-02
/ PRIOR APPLICATION NUMBER: DE 10032529.7
```

```
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: DE 10043826.1
/ PRIOR FILING DATE: 2000-09-01
/ NUMBER OF SEQ ID NOS: 2424
/ SEQ ID NO 1627
/ LENGTH: 5228
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1627

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 6; Length 5228;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
    |||||
Db 336 AAAAAAAAACTATAGCTTTTCTT 312

RESULT 32
US-10-681-773-258
/ Sequence 258, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsuzaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO 258
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-258

Query Match
Best Local Similarity 80.0%; Score 20; DB 7; Length 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AAAAACTATAGCTTGATCTT 25
    |||||
Db 1 AAAAACTATAGCTTGATCTT 20

RESULT 33
US-10-425-114-14574
/ Sequence 14574, Application US/10425114
/ Publication No. US20040034888A1
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jingdong
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Screen, Steven E
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 14574
```

LENGTH: 1666
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: Clone ID: LIB23-006-F9_FLI
US-10-425-114-14574

Query Match 79.2%; Score 19.8; DB 7; Length 1666;
Best Local Similarity 91.3%; Pred. No. 8.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAAAAAGTATAGCTTGATCTT 25
|||||

Db 1643 AAAAAAAGTATAGCTTGATCTT 1665

RESULT 34
US-10-425-115-86064
Sequence 86064, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 86064
LENGTH: 264
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_1784C.1
US-10-425-115-86064

Query Match 77.6%; Score 19.4; DB 8; Length 264;
Best Local Similarity 95.2%; Pred. No. 9.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 AAAAAAATAGCTTGATCTT 25
|||||

Db 221 AAAAAAATAGCTTGATCTT 241

RESULT 35
US-10-087-192-754
Sequence 754, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 754
LENGTH: 276276
TYPE: DNA
ORGANISM: Homo sapiens
US-10-087-192-754

Query Match 77.6%; Score 19.4; DB 5; Length 276276;

Best Local Similarity 95.2%; Pred. No. 2.7e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAGTATAGCTTGA 21
|||||

Db 17567 AAAAAAAGTATAGCTTGA 17587

RESULT 36
US-10-437-963-42936
Sequence 42936, Application US/10437963
Publication No. US2004012343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 42936
LENGTH: 335
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_46140C.1
US-10-437-963-42936

Query Match 76.8%; Score 19.2; DB 7; Length 335;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAGTATAGCTTGATCTT 25
|||||

Db 41 AAAAAAAGTATAGCTTGATCTT 64

RESULT 37
US-10-425-115-144586
Sequence 144586, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 144586
LENGTH: 405
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_63348C.1
US-10-425-115-144586

Query Match 76.8%; Score 19.2; DB 8; Length 405;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAGTATAGCTTGATCT 24
|||||

```
Db          3 AAAAAAAAAAATGCTTGATCT 26

RESULT 38
US-10-915-740A-107/c
; Sequence 107, Application US/10915740A
; Publication No. US20050191316A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tetelien, Heire
; APPLICANT: Venter, J. Craig
; APPLICANT: Maignani, Vega
; APPLICANT: Galeotti, Cecilia
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scariato, Maria
; APPLICANT: Scariato, Vincenzo
; APPLICANT: Rappelli, Rino
; APPLICANT: Piza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: 002441.00090
; CURRENT APPLICATION NUMBER: US/10/915,740A
; CURRENT FILING DATE: 2004-08-11
; PRIOR APPLICATION NUMBER: 09/806,866
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 107
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-915-740A-107

Query Match          76.8%; Score 19.2; DB 9; Length 467;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY          1 AAAAAAAAAAATGCTTGATCT 24
Db          233 AAAAAAAAAAATGCTTGATCT 210

RESULT 39
US-10-972-079-65854/c
; Sequence 65854, Application US/10972079
; Publication No. US2005015317A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65854
; LENGTH: 469

; TYPE: DNA
; ORGANISM: Chicken 19866894333450_3
US-10-972-079-65854

Query Match          76.8%; Score 19.2; DB 9; Length 469;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY          1 AAAAAAAAAAATGCTTGATCT 24
Db          359 AAAAAAAAAAATGCTTGATTT 336

RESULT 40
US-09-925-065A-334939/c
; Sequence 334939, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334939
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-334939

Query Match          76.8%; Score 19.2; DB 4; Length 521;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY          1 AAAAAAAAAAATGCTTGATCT 24
Db          38 AAAAAAAAAAATGCTTTATCT 15

RESULT 41
US-10-369-493-30132/c
; Sequence 30132, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 30132
; LENGTH: 541
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
```

US-10-369-493-30132

Query Match 76.8%; Score 19.2; DB 6; Length 541;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 294 AAAAAAAAAATTATAGATTGATGT 271

RESULT 42

US-09-925-065A-594437
; Sequence 594437, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 594437
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-594437

Query Match 76.8%; Score 19.2; DB 4; Length 552;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 476 AAAAAAAAAAGCTATACATGATCT 499

RESULT 43
US-09-925-065A-264897
; Sequence 264897, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264897

; LENGTH: 618

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-264897

Query Match 76.8%; Score 19.2; DB 4; Length 618;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCT 25
Db 80 AAAAAAAAAATTATAGCTTGATCTT 103

RESULT 44

US-09-925-065A-870597/c
; Sequence 870597, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 870597
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-870597

Query Match 76.8%; Score 19.2; DB 4; Length 669;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 626 AAAAAAAAAAGCTATACATGATCT 603

RESULT 45
US-10-027-632-146140/c
; Sequence 146140, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23

```

; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 146140
; LENGTH: 788
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-146140

Query Match          76.8%; Score 19.2; DB 5; Length 788;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCTT 25
Db 747 AAAAAAAAACTTGTGCTTATCTT 724

RESULT 46
US-10-027-632-146140/c
; Sequence 146140, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 146140
; LENGTH: 788
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-146140

Query Match          76.8%; Score 19.2; DB 6; Length 788;
Best Local Similarity 87.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCTT 25
Db 747 AAAAAAAAACTTGTGCTTATCTT 724

RESULT 47
US-10-369-493-29566
; Sequence 29566, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 29566
; LENGTH: 2123
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; US-10-369-493-29566

Query Match          76.8%; Score 19.2; DB 6; Length 2123;
Best Local Similarity 87.5%; Pred. No. 1.6e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 353 AAAAAAAAACTATTTGCTTGTCT 376

RESULT 48
US-09-925-065A-711163/c
; Sequence 711163, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 711163
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-711163

Query Match          76.8%; Score 19.2; DB 4; Length 2733;
Best Local Similarity 87.5%; Pred. No. 1.6e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 1499 AAAAAAAAAAATAGCTTGATGT 1476

RESULT 49
US-11-097-143-10747
; Sequence 10747, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Ventier, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CLO00728
; CURRENT APPLICATION NUMBER: US/11/097,143
```

```

; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10747
; LENGTH: 6141
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-10747

```

```

Query Match      76.8%; Score 19.2; DB 10; Length 6141;
Best Local Similarity 87.5%; Pred. No. 1.8e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      2 AAAAAAAAACTATAGCTTGATCTT 25
Db      4636 AAAAAATAACTATAGTTGATATT 4659

```

```

RESULT 50
US-10-257-166-4/c
; Sequence 4, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; PRIOR FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 4
; LENGTH: 6247
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-4

```

```

Query Match      76.8%; Score 19.2; DB 7; Length 6247;
Best Local Similarity 87.5%; Pred. No. 1.8e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAACTATAGCTTGATCT 24
Db      2715 AAAAAATAACTATAGCTTGATCT 2692

```

Search completed: December 14, 2005, 08:46:21
Job time : 376.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 ; Search time 373.2 Seconds
(without alignments) 553.951 Million cell updates/sec

Title: US-10-681-773-5

Perfect score: 25
Sequence: 1 aaaaaaaaaagcatgactgtgacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Published Applications NA Main.*

1: /cgn2_6/ptodaca/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodaca/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodaca/1/pubpna/US09_PUBCOMB.seq:*
4: /cgn2_6/ptodaca/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodaca/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodaca/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodaca/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodaca/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodaca/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodaca/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	7	US-10-681-773-5
2	24.6	98.4	33	9	US-10-891-260-3906
3	24.6	98.4	515	5	US-10-027-632-66793
4	24.6	98.4	515	5	US-10-027-632-66794
5	24.6	98.4	515	6	US-10-027-632-66793
6	24.6	98.4	515	6	US-10-027-632-66794
7	24.6	98.4	651	5	US-10-027-632-34604
8	24.6	98.4	651	5	US-10-027-632-34605
9	24.6	98.4	651	6	US-10-027-632-34604
10	24.6	98.4	651	6	US-10-027-632-34605
11	23.4	93.6	25	7	US-10-681-773-6
12	22.4	89.6	2013	3	US-09-938-842A-1519
13	22.4	89.6	2013	3	US-09-938-842A-1519
14	22	88.0	25	7	US-10-681-773-41
15	21	84.0	25	7	US-10-681-773-113
16	20.4	81.6	25	7	US-10-681-773-42
17	20.2	80.8	161	7	US-10-424-599-88653
18	20.2	80.8	599	9	US-10-972-079-50639
19	20.2	80.8	1066	7	US-10-437-963-10167
20	20	79.2	25	7	US-10-681-773-315
21	19.8	78.0	525	4	US-09-925-065A-302662
22	19.8	79.2	217409	5	US-10-087-192-1954
23	19.4	77.6	25	7	US-10-681-773-115

24	19.4	77.6	609	4	US-09-925-065A-436983	Sequence 436983, App
25	19.4	77.6	32404	3	US-09-997-722-160	Sequence 160, App
26	19.2	76.8	321	7	US-10-424-599-52989	Sequence 52989, A
27	19.2	76.8	5716	4	US-09-925-065A-387162	Sequence 387162, App
28	19.2	76.8	594	4	US-09-925-065A-948817	Sequence 948817, App
29	19.2	76.8	594	4	US-09-925-065A-948818	Sequence 948818, App
30	19.2	76.8	606	4	US-09-925-065A-888192	Sequence 888192, App
31	19.2	76.8	634	4	US-09-925-065A-887791	Sequence 887791, App
32	19.2	76.8	659	4	US-09-925-065A-883697	Sequence 883697, App
33	19.2	76.8	659	4	US-09-925-065A-918097	Sequence 918097, App
34	19.2	76.8	756	8	US-10-425-115-80855	Sequence 80855, A
35	19.2	76.8	795	4	US-09-925-065A-34898	Sequence 34898, A
36	19.2	76.8	2775	4	US-09-925-065A-720587	Sequence 720587, App
37	19.2	76.8	3956	8	US-10-723-860-5662	Sequence 5662, App
38	19.2	76.8	7764	3	US-09-914-353-21548	Sequence 21548, A
39	19.2	76.8	7764	6	US-10-941-434-60	Sequence 60, App1
40	19.2	76.8	7764	8	US-10-651-237-27	Sequence 27, App1
41	19.2	76.8	7764	8	US-10-782-413-27	Sequence 27, App1
42	19.2	76.8	7764	9	US-10-934-998-33	Sequence 33, App1
43	19.2	76.8	7764	9	US-10-479-874A-3	Sequence 3, App1
44	19.2	76.8	130030	8	US-10-719-993-6986	Sequence 6986, App
45	19.2	76.8	193757	8	US-10-719-993-6939	Sequence 6939, App
46	19.2	76.8	196151	9	US-10-981-277-51	Sequence 51, App1
47	19.2	76.8	1790342	8	US-10-719-993-6840	Sequence 6840, App
48	18.8	75.2	201	8	US-10-719-993-16941	Sequence 16941, A
49	18.8	75.2	494	4	US-09-925-065A-114553	Sequence 114553, A
50	18.8	75.2	545	5	US-10-027-632-88766	Sequence 88766, A
51	18.8	75.2	545	5	US-10-027-632-88767	Sequence 88767, A
52	18.8	75.2	545	6	US-10-027-632-88766	Sequence 88766, A
53	18.8	75.2	545	6	US-10-027-632-88767	Sequence 88767, A
54	18.8	75.2	587	4	US-09-925-065A-844286	Sequence 844286, App
55	18.8	75.2	594	4	US-09-925-065A-926372	Sequence 926372, App
56	18.8	75.2	644	5	US-10-027-632-85415	Sequence 85415, A
57	18.8	75.2	644	5	US-10-027-632-109515	Sequence 109515, A
58	18.8	75.2	644	6	US-10-027-632-85415	Sequence 85415, A
59	18.8	75.2	644	6	US-10-027-632-109515	Sequence 109515, A
60	18.8	75.2	781	7	US-10-424-599-81767	Sequence 81767, A
61	18.8	75.2	905	4	US-09-925-065A-90459	Sequence 90459, A
62	18.8	75.2	1659	7	US-10-425-114-29204	Sequence 29204, A
63	18.8	75.2	2613	10	US-11-097-143-23146	Sequence 23146, A
64	18.8	75.2	44063	7	US-10-322-281-718	Sequence 718, App
65	18.8	75.2	81460	7	US-10-451-467A-659	Sequence 659, App
66	18.8	75.2	366803	8	US-10-719-993-6805	Sequence 6805, App
67	18.8	75.2	513509	3	US-09-754-853A-4	Sequence 4, App1
68	18.6	74.4	249	7	US-10-424-599-50146	Sequence 50146, A
69	18.6	74.4	361	8	US-10-357-930-1004	Sequence 1004, App
70	18.6	74.4	396	8	US-10-357-930-10173	Sequence 10173, A
71	18.6	74.4	415	8	US-10-357-930-31371	Sequence 31371, A
72	18.6	74.4	415	8	US-10-357-930-40341	Sequence 40341, A
73	18.6	74.4	415	8	US-10-357-930-40341	Sequence 40341, A
74	18.6	74.4	501	7	US-10-424-599-91078	Sequence 91078, A
75	18.6	74.4	505	5	US-10-027-632-216787	Sequence 216787, App
76	18.6	74.4	505	6	US-10-027-632-216787	Sequence 216787, App
77	18.6	74.4	505	6	US-10-027-632-216787	Sequence 216787, App
78	18.6	74.4	517	7	US-10-437-963-46104	Sequence 46104, A
79	18.6	74.4	540	4	US-09-925-065A-474718	Sequence 474718, App
80	18.6	74.4	561	7	US-10-021-323-13985	Sequence 13985, A
81	18.6	74.4	570	4	US-09-925-065A-261160	Sequence 261160, App
82	18.6	74.4	580	4	US-09-925-065A-261161	Sequence 261161, App
83	18.6	74.4	586	4	US-09-925-065A-423130	Sequence 423130, App
84	18.6	74.4	586	4	US-09-925-065A-779600	Sequence 779600, App
85	18.6	74.4	586	4	US-09-925-065A-779601	Sequence 779601, App
86	18.6	74.4	586	4	US-09-925-065A-840920	Sequence 840921, App
87	18.6	74.4	594	5	US-10-027-632-137080	Sequence 137080, App
88	18.6	74.4	594	5	US-10-027-632-137081	Sequence 137081, App
89	18.6	74.4	594	6	US-10-027-632-137080	Sequence 137080, App
90	18.6	74.4	594	6	US-10-027-632-137081	Sequence 137081, App
91	18.6	74.4	594	6	US-10-027-632-137082	Sequence 137082, App
92	18.6	74.4	594	6	US-10-027-632-137082	Sequence 137082, App
93	18.6	74.4	595	4	US-09-925-065A-710088	Sequence 710088, App
94	18.6	74.4	600	9	US-10-972-079-51006	Sequence 51006, A
95	18.6	74.4	600	9	US-10-972-079-51008	Sequence 51008, A
96	18.6	74.4	600	9	US-10-972-079-51009	Sequence 51009, A

```
97 18.6 74.4 600 9 US-10-922-079-79144 Sequence 79144, A
98 18.6 74.4 647 4 US-09-923-065A-335948 Sequence 335948, A
99 18.6 74.4 686 8 US-10-357-930-20656 Sequence 20656, A
100 18.6 74.4 686 8 US-10-357-930-22894 Sequence 22894, A
101 18.6 74.4 686 8 US-10-357-930-26496 Sequence 26496, A
102 18.6 74.4 686 8 US-10-357-930-28753 Sequence 28753, A
103 18.6 74.4 711 8 US-10-723-860-8049 Sequence 8049, A
104 18.6 74.4 725 8 US-10-423-115-102172 Sequence 102172, A
105 18.6 74.4 732 4 US-09-923-065A-933507 Sequence 933507, A
106 18.6 74.4 732 4 US-09-923-065A-952072 Sequence 952072, A
107 18.6 74.4 779 5 US-10-027-632-143224 Sequence 143224, A
108 18.6 74.4 779 5 US-10-027-632-143225 Sequence 143225, A
109 18.6 74.4 779 6 US-10-027-632-143224 Sequence 143224, A
110 18.6 74.4 779 6 US-10-027-632-143225 Sequence 143225, A
111 18.6 74.4 894 4 US-09-923-065A-91822 Sequence 91822, A
112 18.6 74.4 894 4 US-09-923-065A-91823 Sequence 91823, A
113 18.6 74.4 894 4 US-09-923-065A-91824 Sequence 91824, A
114 18.6 74.4 894 4 US-09-923-065A-91825 Sequence 91825, A
115 18.6 74.4 894 4 US-09-923-065A-91826 Sequence 91826, A
116 18.6 74.4 898 7 US-10-424-599-3082 Sequence 3082, A
117 18.6 74.4 1168 4 US-09-923-065A-711639 Sequence 711639, A
118 18.6 74.4 1168 4 US-09-923-065A-711640 Sequence 711640, A
119 18.6 74.4 1168 4 US-09-923-065A-711641 Sequence 711641, A
120 18.6 74.4 1168 4 US-09-923-065A-711642 Sequence 711642, A
121 18.6 74.4 1197 8 US-10-474-792-295 Sequence 295, A
122 18.6 74.4 1268 7 US-10-423-114-7757 Sequence 7757, A
123 18.6 74.4 1670 4 US-09-923-065A-34207 Sequence 34207, A
124 18.6 74.4 1670 4 US-09-923-065A-65780 Sequence 65780, A
125 18.6 74.4 1706 4 US-09-923-065A-61913 Sequence 61913, A
126 18.6 74.4 1706 4 US-09-923-065A-61914 Sequence 61914, A
127 18.6 74.4 1706 4 US-09-923-065A-61915 Sequence 61915, A
128 18.6 74.4 1706 4 US-09-923-065A-61916 Sequence 61916, A
129 18.6 74.4 1718 7 US-10-424-599-127935 Sequence 127935, A
130 18.6 74.4 6495 3 US-09-764-891-9129 Sequence 9129, A
131 18.6 74.4 6495 3 US-10-091-572-705 Sequence 705, A
132 18.6 74.4 14103 6 US-10-242-355-1114 Sequence 1114, A
133 18.6 74.4 14779 10 US-11-097-143-37828 Sequence 37828, A
134 18.6 74.4 25032 7 US-10-394-948-19 Sequence 19, A
135 18.6 74.4 25032 7 US-10-052-482-1 Sequence 1, A
136 18.6 74.4 59065 6 US-10-135-696-3 Sequence 3, A
137 18.6 74.4 59065 8 US-10-820-230-3 Sequence 3, A
138 18.6 74.4 60815 5 US-10-087-192-52 Sequence 52, A
139 18.6 74.4 128978 8 US-10-775-169-345 Sequence 345, A
140 18.6 74.4 160552 8 US-10-697-828-11 Sequence 11, A
141 18.6 74.4 163321 5 US-10-087-192-76 Sequence 76, A
142 18.6 74.4 164841 5 US-10-087-192-871 Sequence 871, A
143 18.6 74.4 276276 5 US-10-087-192-754 Sequence 754, A
144 18.4 73.6 25 7 US-10-681-773-316 Sequence 316, A
145 18.4 73.6 479 4 US-09-923-065A-870488 Sequence 870488, A
146 18.4 73.6 587 4 US-09-923-065A-784276 Sequence 784276, A
147 18.4 73.6 599 4 US-09-923-065A-388264 Sequence 388264, A
148 18.4 73.6 627 4 US-09-923-065A-874821 Sequence 874821, A
149 18.4 73.6 1298 7 US-10-425-114-13224 Sequence 13224, A
150 18.4 73.6 2159 7 US-10-424-599-84242 Sequence 84242, A
```

ALIGNMENTS

```
RESULT 1
US-10-681-773-5
; Sequence 5, Application US/10681773
; Publication No. US2004016890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO: 5
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-5
```

```
Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 1 AAAAAAAAAAGCATGCTGTGACAC 25
```

```
RESULT 2
US-10-891-260-3906
; Sequence 3906, Application US/10891260
; Publication No. US20050227244A1
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc.
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; TITLE OF INVENTION: Methods of Analysis of Human Polymorphisms
; FILE REFERENCE: 3522.3
; CURRENT APPLICATION NUMBER: US/10/891,260
; PRIOR FILING DATE: 2004-07-13
; PRIOR APPLICATION NUMBER: 10/681,773
; PRIOR FILING DATE: 2003-10-07
; NUMBER OF SEQ ID NOS: 10244
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 3906
; LENGTH: 33
; TYPE: DNA
; ORGANISM: homo sapien
US-10-891-260-3906
```

```
Query Match 98.4%; Score 24.6; DB 9; Length 33;
Best Local Similarity 96.0%; Pred. No. 1.7;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 1 AAAAAAAAAAGCATGCTGTGACAC 25
```

```
RESULT 3
US-10-027-632-66793
; Sequence 66793, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
```

```
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 66793
;; LENGTH: 515
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-66793
```

```
Query Match          98.4%: Score 24.6; DB 5; Length 515;
Best Local Similarity 96.0%: Pred. No. 2.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGACTGTGACAC 32
```

```
RESULT 4
US-10-027-632-66794
; Sequence 66794, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66794
; LENGTH: 515
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-66794

Query Match          98.4%: Score 24.6; DB 5; Length 515;
Best Local Similarity 96.0%: Pred. No. 2.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGACTGTGACAC 32
```

```
RESULT 5
US-10-027-632-66793
; Sequence 66793, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```

```
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 66793
;; LENGTH: 515
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-66793
```

```
Query Match          98.4%: Score 24.6; DB 6; Length 515;
Best Local Similarity 96.0%: Pred. No. 2.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGACTGTGACAC 32
```

```
RESULT 6
US-10-027-632-66794
; Sequence 66794, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66794
; LENGTH: 515
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-66794
```

```
Query Match          98.4%: Score 24.6; DB 6; Length 515;
Best Local Similarity 96.0%: Pred. No. 2.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGACTGTGACAC 32
```

```
RESULT 7
US-10-027-632-34604
; Sequence 34604, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34604
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34604
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 651;
Best Local Similarity 96.0%; Pred. No. 2.7;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGACTGTGCAC 25
Db 169 AAAAAAAAAAGCATGATGTGCAC 193
```

```
RESULT 8
US-10-027-632-34605
; Sequence 34605, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34605
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Human
```

```
US-10-027-632-34605
Query Match          98.4%; Score 24.6; DB 5; Length 651;
Best Local Similarity 96.0%; Pred. No. 2.7;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGACTGTGCAC 25
Db 169 AAAAAAAAAAGCATGATGTGCAC 193
```

```
RESULT 9
US-10-027-632-34604
; Sequence 34604, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34604
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34604
```

```
Query Match          98.4%; Score 24.6; DB 6; Length 651;
Best Local Similarity 96.0%; Pred. No. 2.7;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGACTGTGCAC 25
Db 169 AAAAAAAAAAGCATGATGTGCAC 193
```

```
RESULT 10
US-10-027-632-34605
; Sequence 34605, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34605
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34605
```

Query Match 98.4%; Score 24.6; DB 6; Length 651;
Best Local Similarity 96.0%; Pred. No. 2.7;

Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 169 AAAAAAAAAAGCATGACTGTGACAC 193
```

```

RESULT 11
; Sequence 6, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 6
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-6
```

Query Match 93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 5;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 1 AAAAAAAAAAGCATGACTGTGACAC 25
```

```

RESULT 12
US-09-938-842A-1519/c
; Sequence 1519, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
```

```

; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1519
; LENGTH: 2013
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1519
```

Query Match 89.6%; Score 22.4; DB 3; Length 2013;
Best Local Similarity 95.8%; Pred. No. 26;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 1329 AAAAAAAAAAGCATGACTGTGACA 1306
```

```

RESULT 13
US-09-938-842A-1519/c
; Sequence 1519, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1519
; LENGTH: 2013
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1519
```

Query Match 89.6%; Score 22.4; DB 3; Length 2013;
Best Local Similarity 95.8%; Pred. No. 26;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 1329 AAAAAAAAAAGCATGACTGTGACA 1306
```

```

RESULT 14
US-10-681-773-41
; Sequence 41, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
```

```

; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 41
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-41

Query Match      88.0%; Score 22; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AAAAAAGCATGCTGTGACAC 25
Db 1 AAAAAAGCATGCTGTGACAC 22

RESULT 15
US-10-681-773-113
; Sequence 113, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 113
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-113

Query Match      84.0%; Score 21; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 AAAAAAGCATGCTGTGACAC 25
Db 1 AAAAAAGCATGCTGTGACAC 21

RESULT 16
US-10-681-773-42
; Sequence 42, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 42
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien

US-10-681-773-42

Query Match      81.6%; Score 20.4; DB 7; Length 25;
Best Local Similarity 95.5%; Pred. No. 89;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAAAAAGCATGCTGTGACAC 25
Db 1 AAAAAAGCATGCTGTGACAC 22

RESULT 17
US-10-424-599-88693
; Sequence 88693, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 88693
; LENGTH: 161
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51098C.1
US-10-424-599-88693

Query Match      80.8%; Score 20.2; DB 7; Length 161;
Best Local Similarity 88.0%; Pred. No. 1,5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAGCATGCTGTGACAC 25
Db 95 AAAAAAGCATGCTGTGACAC 119

RESULT 18
US-10-972-079-50639/c
; Sequence 50639, Application US/10972079
; Publication No. US2005015317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEST
; FILE REFERENCE: MM11110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 50639
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Chicken 19866894297718_1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(193)
; OTHER INFORMATION: n is any nucleotide
US-10-972-079-50639
```

Query Match 80.8%; Score 20.2; DB 9; Length 599;
Best Local Similarity 88.0%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGCAC 25

Db 490 AAAAAAAAAAGCATGACTGTTCGCC 466

RESULT 19

US-10-437-963-101667/c
; Sequence 101667, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 101667
; LENGTH: 1066
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_99266C.1
US-10-437-963-101667

Query Match 80.8%; Score 20.2; DB 7; Length 1066;
Best Local Similarity 88.0%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGCAC 25

Db 788 AAAAAAAAAAGCATGACTGTGCAC 764

RESULT 20

US-10-681-773-315
; Sequence 315, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 315
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-315

Query Match 80.0%; Score 20; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AAAAAAAAAAGCATGACTGTGCAC 25

Db 1 AAAAAAAAAAGCATGACTGTGCAC 20

RESULT 21

US-09-925-065A-302662
; Sequence 302662, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957066
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 302662
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-302662

Query Match 79.2%; Score 19.8; DB 4; Length 525;
Best Local Similarity 91.3%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGACTGTGACA 24

Db 190 AAAAAAAAAAGCATGACTGTGACA 212

RESULT 22

US-10-087-192-1954
; Sequence 1954, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1954
; LENGTH: 217409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(217409)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1954

Query Match 79.2%; Score 19.8; DB 5; Length 217409;


```
RESULT 27
US-09-925-065A-387162
; Sequence 387162, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 387162
; LENGTH: 571
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-387162

Query Match      76.8%; Score 19.2; DB 4; Length 571;
Best Local Similarity 87.5%; Pred. No. 4.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 77 AAAAAAAAAAGCAAGCTCTGACA 100

RESULT 28
US-09-925-065A-948817/c
; Sequence 948817, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 948817
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-948817

Query Match      76.8%; Score 19.2; DB 4; Length 594;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 162 AGAAAAAAAAAGCAAGCTGTGACA 139

RESULT 29
US-09-925-065A-948818/c
; Sequence 948818, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 948818
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-948818

Query Match      76.8%; Score 19.2; DB 4; Length 594;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 162 AGAAAAAAAAAGCAAGCTGTGACA 139

RESULT 30
US-09-925-065A-888192/c
; Sequence 888192, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 888192
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-888192

Query Match      76.8%; Score 19.2; DB 4; Length 606;
```

Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
|||||
Db 56 AAAAAAAAAAGCATGACTGTGACA 33

RESULT 31

US-09-925-065A-887791/c
; Sequence 887791, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887791
; LENGTH: 634
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-887791

Query Match 76.8%; Score 19.2; DB 4; Length 634;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
|||||
Db 59 AAAAAAAAAAGCATGACTGTGACA 36

RESULT 32

US-09-925-065A-883697/c
; Sequence 883697, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 883697
; LENGTH: 646
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-925-065A-883697

Query Match 76.8%; Score 19.2; DB 4; Length 646;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
|||||
Db 121 AAAAAAAAAAGCATGCTGGGCA 98

RESULT 33

US-09-925-065A-918097
; Sequence 918097, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 918097
; LENGTH: 659
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-918097

Query Match 76.8%; Score 19.2; DB 4; Length 659;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
|||||
Db 73 AAAAAAAAAAGCATGACTGTGACA 96

RESULT 34

US-10-425-115-80855/c
; Sequence 80855, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 80855
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: MRT4577_173758C.1
US-10-425-115-80855

Query Match 76.8%; Score 19.2; DB 8; Length 756;

Best Local Similarity 87.5%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGCTGTGACAC 25
Db 92 AAAAAAAAAATCATCTCTGTATACAC 69

RESULT 35
US-09-925-065A-34898

; Sequence 34898, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A

; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34898
; LENGTH: 795
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-925-065A-34898

Query Match 76.8%; Score 19.2; DB 4; Length 795;
Best Local Similarity 87.5%; Pred. No. 4.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGCTGTGACAC 25
Db 221 AAAAAAAAAAGAAATGTGACAC 244

RESULT 36

US-09-925-065A-720587/C
; Sequence 720587, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A

; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 720587
; LENGTH: 2775
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-925-065A-720587

Query Match 76.8%; Score 19.2; DB 4; Length 2775;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
Db 1373 AAAAAAAAAATCGTCCTGTGACA 1350

RESULT 37

US-10-723-860-5662
; Sequence 5662, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:

; APPLICANT: Aziz, Nafasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5662
; LENGTH: 3956
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3825) .. (3853)
; OTHER INFORMATION: n is a, c, g, or t

US-10-723-860-5662

Query Match 76.8%; Score 19.2; DB 8; Length 3956;
Best Local Similarity 87.5%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
Db 3298 AAAAAAAAAAGCATATCTGTGACA 3321

RESULT 38

US-09-814-353-21548/C
; Sequence 21548, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:

; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-0068
; CURRENT APPLICATION NUMBER: US/09/814,353
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21

```

; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21548
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3, 4, 5, 6, 7, 7755, 7756, 7757, 7758, 7759, 7760,
; LOCATION: 7761, 7762, 7763, 7764
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21548

Query Match          76.8%; Score 19.2; DB 3; Length 7764;
Best Local Similarity 87.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGACTGTGACAC 25
DB 6509 AAAAAAAGCAGCAGACTTTGACAC 6486

RESULT 39
US-10-341-434-60/C
; Sequence 60, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 60
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (708)..(5924)
; OTHER INFORMATION:
US-10-341-434-60

Query Match          76.8%; Score 19.2; DB 6; Length 7764;
Best Local Similarity 87.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGACTGTGACAC 25
DB 6509 AAAAAAAGCAGCAGACTTTGACAC 6486

RESULT 40
US-10-651-237-27/C
; Sequence 27, Application US/10651237
; Publication No. US20050048494A1
; GENERAL INFORMATION:
; APPLICANT: Ortho-Clinical Diagnostics, Inc.
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: Colorectal Cancer Prognostics
; FILE REFERENCE: ADS-5003 US NP
; CURRENT APPLICATION NUMBER: US/10/651,237
; CURRENT FILING DATE: 2003-08-27
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 27
; LENGTH: 7764
; TYPE: DNA
```

```

; ORGANISM: human
US-10-651-237-27

Query Match          76.8%; Score 19.2; DB 8; Length 7764;
Best Local Similarity 87.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGACTGTGACAC 25
DB 6509 AAAAAAAGCAGCAGACTTTGACAC 6486

RESULT 41
US-10-782-413-27/C
; Sequence 27, Application US/10782413
; Publication No. US20050048526A1
; GENERAL INFORMATION:
; APPLICANT: Ortho-Clinical Diagnostics, Inc.
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: Colorectal Cancer Prognostics
; FILE REFERENCE: VDX-5002 CIP
; CURRENT APPLICATION NUMBER: US/10/782,413
; CURRENT FILING DATE: 2004-02-18
; PRIOR APPLICATION NUMBER: 10/651,237
; PRIOR FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 27
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: human
US-10-782-413-27

Query Match          76.8%; Score 19.2; DB 8; Length 7764;
Best Local Similarity 87.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGACTGTGACAC 25
DB 6509 AAAAAAAGCAGCAGACTTTGACAC 6486

RESULT 42
US-10-934-998-33/C
; Sequence 33, Application US/10934998
; Publication No. US20050153917A1
; GENERAL INFORMATION:
; APPLICANT: AL-MAHMOOD, SALMAN
; APPLICANT: SCHNEIDER, CHRISTOPHE
; TITLE OF INVENTION: GENES INVOLVED IN REGULATING ANGIOGENESIS, PHARMACEUTICAL
; FILE REFERENCE: BMA-04-1206
; CURRENT APPLICATION NUMBER: US/10/934,998
; CURRENT FILING DATE: 2004-09-03
; PRIOR APPLICATION NUMBER: PCT/FR03/00695
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: FR02/02717
; PRIOR FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: FR02/04546
; PRIOR FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 33
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: GS-N32
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GENBANK/NM_001271
; DATABASE ENTRY DATE: 2000-10-31
; RELEVANT RESIDUES: (1)..(7764)
```

US-10-934-998-33

Query Match 76.8%; Score 19.2; DB 9; Length 7764;
Best Local Similarity 87.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGCTGTGACAC 25
DB 6509 AAAAAAAGCAGCAGCTTTGACAC 6486

RESULT 43

US-10-479-874A-3/C
; Sequence 3, Application US/10479874A
; Publication No. US20050170344A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: CHDS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-090C-US
; CURRENT APPLICATION NUMBER: US/10/479,874A
; CURRENT FILING DATE: 2003-12-04
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/338,733
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/357,600
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 7764
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-479-874A-3

Query Match 76.8%; Score 19.2; DB 9; Length 7764;
Best Local Similarity 87.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGCTGTGACAC 25
DB 6509 AAAAAAAGCAGCAGCTTTGACAC 6486

RESULT 44

US-10-719-993-6986/C
; Sequence 6986, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6986
; LENGTH: 130030
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(130030)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6986

Query Match 76.8%; Score 19.2; DB 8; Length 130030;
Best Local Similarity 87.5%; Pred. No. 1.1e+03;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
DB 21508 AAAAAAAAAAGCATGCTGTGTCA 21485

RESULT 45

US-10-719-993-6939
; Sequence 6939, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6939
; LENGTH: 193757
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(193757)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6939

Query Match 76.8%; Score 19.2; DB 8; Length 193757;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
DB 85989 AAAAAAAAAAGCATGTTGTGTCA 86012

RESULT 46

US-10-981-277-51
; Sequence 51, Application US/10981277
; Publication No. US20050181389A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa
; TITLE OF INVENTION: Compositions and Methods for Glioma Classification
; FILE REFERENCE: 03-968-US
; CURRENT APPLICATION NUMBER: US/10/981,277
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/516,817
; PRIOR FILING DATE: 2003-11-03
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 51
; LENGTH: 196151
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-981-277-51

Query Match 76.8%; Score 19.2; DB 9; Length 196151;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
DB 7122 AAAAAAAAAAGACTGCTGAGACA 7145

RESULT 47

US-10-719-993-6940/C
; Sequence 6940, Application US/10719993
; Publication No. US20040265849A1

```
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CL001496
/ CURRENT APPLICATION NUMBER: US/10/719,993
/ NUMBER OF SEQ ID NOS: 55342
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 6940
/ LENGTH: 1790242
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURES:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(1790242)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6940

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 8; Length 1790242;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 705163 AAAAAAAAAAGCATGTTGTGTCA 705140

RESULT 48
US-10-719-993-16941/C
/ Sequence 16941, Application US/10719993
/ Publication No. US20040265849A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CL001496
/ CURRENT APPLICATION NUMBER: US/10/719,993
/ NUMBER OF SEQ ID NOS: 55342
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 16941
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-719-993-16941

Query Match
Best Local Similarity 75.2%; Score 18.8; DB 8; Length 201;
Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 109 AAGAAAAAAMTGTATGACTGTGACA 86

RESULT 49
US-09-925-065A-114553/C
/ Sequence 114553, Application US/09925065A
/ Publication No. US20050228172A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ FILE REFERENCE: 108827.135
/ CURRENT APPLICATION NUMBER: US/09/925,065A
/ CURRENT FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243,096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
```

```
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289,846
/ PRIOR FILING DATE: 2001-05-09
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 114553
/ LENGTH: 494
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-925-065A-114553

Query Match
Best Local Similarity 75.2%; Score 18.8; DB 4; Length 494;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGCATGACTGTGACA 24
DB 407 AAAAAAAAAATCATGACTGTAA 386

RESULT 50
US-10-027-632-88766/C
/ Sequence 88766, Application US/10027632
/ Publication No. US20020198371A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 88766
/ LENGTH: 545
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-88766

Query Match
Best Local Similarity 90.9%; Score 18.8; DB 5; Length 545;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
DB 154 AAAAAAAAAAGCATGCTGTAA 133

Search completed: December 14, 2005, 08:46:27
Job time : 379.2 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 ; Search time 373.2 Seconds
(without alignments)

553.951 Million cell updates/sec

Title: US-10-681-773-6

Perfect score: 25
Sequence: 1 aaaaaaaaaagcatgtatgtgacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications NA Main: *
1: /cgn2_6/ptodata/1/pubpna/us07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/1/pubpna/us08_PUBCOMB.seq: *
3: /cgn2_6/ptodata/1/pubpna/us09_PUBCOMB.seq: *
4: /cgn2_6/ptodata/1/pubpna/us09B_PUBCOMB.seq: *
5: /cgn2_6/ptodata/1/pubpna/us10A_PUBCOMB.seq: *
6: /cgn2_6/ptodata/1/pubpna/us10C_PUBCOMB.seq: *
7: /cgn2_6/ptodata/1/pubpna/us10D_PUBCOMB.seq: *
8: /cgn2_6/ptodata/1/pubpna/us10E_PUBCOMB.seq: *
9: /cgn2_6/ptodata/1/pubpna/us11_PUBCOMB.seq: *
10: /cgn2_6/ptodata/1/pubpna/us11_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	7	US-10-681-773-6
2	24.6	98.4	33	9	US-10-681-773-6
3	24.6	98.4	515	5	US-10-027-632-66793
4	24.6	98.4	515	5	US-10-027-632-66794
5	24.6	98.4	515	6	US-10-027-632-66793
6	24.6	98.4	515	6	US-10-027-632-66794
7	24.6	98.4	651	5	US-10-027-632-34604
8	24.6	98.4	651	5	US-10-027-632-34605
9	24.6	98.4	651	6	US-10-027-632-34604
10	24.6	98.4	651	6	US-10-027-632-34605
11	23.4	93.6	25	7	US-10-681-773-5
12	22	88.0	25	7	US-10-681-773-42
13	21	84.0	25	7	US-10-681-773-115
14	21	84.0	32404	3	US-09-997-722-160
15	20.8	83.2	2013	3	US-09-938-842A-1519
16	20.8	83.2	2013	3	US-09-938-842A-1519
17	20.8	83.2	130030	8	US-10-719-993-6986
18	20.8	83.2	130757	8	US-10-719-993-6986
19	20.8	83.2	1790242	8	US-10-719-993-6993
20	20.4	81.6	25	7	US-10-681-773-41
21	20.2	80.8	725	8	US-10-425-115-102172
22	20.2	80.8	25032	7	US-10-394-948-19
23	20.2	80.8	25032	7	US-10-052-482-1

24	20.2	80.8	160552	8	US-10-697-828-11	Sequence 11, Appl
25	20	80.0	25	7	US-10-681-773-316	Sequence 316, App
26	19.8	79.2	279	8	US-10-425-115-16636	Sequence 16636, A
27	19.8	79.2	788	4	US-09-925-065A-706877	Sequence 706877, A
28	19.8	79.2	177380	8	US-10-484-577-683	Sequence 683, App
29	19.8	79.2	493631	5	US-10-087-192-205	Sequence 205, App
30	19.4	77.6	25	7	US-10-681-773-113	Sequence 113, App
31	19.4	77.6	329	3	US-09-867-701-4792	Sequence 4792, App
32	19.4	77.6	356	3	US-09-867-701-4792	Sequence 507, App
33	19.4	77.6	356	3	US-09-867-701-4792	Sequence 2965, App
34	19.4	77.6	356	9	US-10-843-641A-2966	Sequence 10644, A
35	19.4	77.6	377	3	US-09-867-701-10644	Sequence 4991, App
36	19.4	77.6	377	3	US-09-867-701-10644	Sequence 10729, A
37	19.4	77.6	366	3	US-09-867-701-10729	Sequence 3652, App
38	19.4	77.6	643	9	US-10-956-157-2976	Sequence 8211, App
39	19.4	77.6	643	9	US-10-956-157-8211	Sequence 143, App
40	19.4	77.6	3434	5	US-10-175-523-143	Sequence 143, App
41	19.4	77.6	3434	10	US-11-099-266-143	Sequence 143, App
42	19.4	77.6	219352	7	US-10-322-81-45	Sequence 45, Appl
43	19.2	76.8	163	7	US-10-021-323-3648	Sequence 3648, App
44	19.2	76.8	166	7	US-10-021-323-3727	Sequence 3727, App
45	19.2	76.8	412	5	US-10-027-632-300127	Sequence 300127, A
46	19.2	76.8	412	6	US-10-027-632-300127	Sequence 300127, A
47	19.2	76.8	414	3	US-09-764-891-716	Sequence 716, App
48	19.2	76.8	438	5	US-10-027-632-62399	Sequence 62399, A
49	19.2	76.8	464	3	US-09-864-761-4453	Sequence 4453, App
50	19.2	76.8	503	5	US-10-027-632-323452	Sequence 323452, A
51	19.2	76.8	503	6	US-10-027-632-323452	Sequence 323452, A
52	19.2	76.8	516	6	US-10-029-386-10809	Sequence 10809, A
53	19.2	76.8	525	4	US-09-925-065A-139586	Sequence 139586, A
54	19.2	76.8	525	4	US-09-925-065A-139587	Sequence 139587, A
55	19.2	76.8	545	4	US-09-925-065A-616706	Sequence 616706, A
56	19.2	76.8	545	4	US-09-925-065A-616707	Sequence 616707, A
57	19.2	76.8	552	9	US-10-972-079-4874	Sequence 4874, App
58	19.2	76.8	555	4	US-09-925-065A-140254	Sequence 140254, App
59	19.2	76.8	558	7	US-10-021-323-3035	Sequence 3035, App
60	19.2	76.8	576	5	US-10-027-632-274722	Sequence 274722, A
61	19.2	76.8	576	6	US-10-027-632-274722	Sequence 274722, A
62	19.2	76.8	585	5	US-10-027-632-268085	Sequence 268085, A
63	19.2	76.8	585	6	US-10-027-632-268085	Sequence 268085, A
64	19.2	76.8	589	4	US-09-925-065A-517778	Sequence 517778, A
65	19.2	76.8	597	4	US-09-925-065A-271047	Sequence 271047, A
66	19.2	76.8	635	4	US-09-925-065A-613077	Sequence 613077, A
67	19.2	76.8	635	4	US-09-925-065A-956849	Sequence 956849, A
68	19.2	76.8	782	5	US-10-202-193-144	Sequence 144, App
69	19.2	76.8	795	4	US-09-925-065A-34898	Sequence 34898, App
70	19.2	76.8	802	4	US-09-925-065A-943919	Sequence 943919, A
71	19.2	76.8	1405	4	US-09-925-065A-943919	Sequence 943919, A
72	19.2	76.8	1654	8	US-10-739-930-2994	Sequence 2994, App
73	19.2	76.8	8050	3	US-09-874-562-11	Sequence 11, Appl
74	19.2	76.8	8050	3	US-10-756-149-90	Sequence 90, Appl
75	19.2	76.8	133300	8	US-10-322-696-112	Sequence 112, App
76	19.2	76.8	188794	9	US-10-981-277-44	Sequence 43, Appl
77	19.2	76.8	193074	9	US-10-737-082-73	Sequence 73, Appl
78	19.2	76.8	519599	9	US-10-765-790-73	Sequence 15235, A
79	19.2	76.8	600	9	US-10-972-079-15235	Sequence 15235, A
80	19.2	76.8	600	9	US-10-972-079-15236	Sequence 15236, A
81	19.2	76.8	600	9	US-10-972-079-15237	Sequence 15237, A
82	19.2	76.8	600	9	US-10-972-079-15238	Sequence 15238, A
83	19.2	76.8	600	9	US-10-972-079-15239	Sequence 15239, A
84	19.2	76.8	600	9	US-10-972-079-15240	Sequence 15240, A
85	19.2	76.8	600	9	US-10-972-079-15241	Sequence 15241, A
86	19.2	76.8	600	9	US-10-972-079-15242	Sequence 15242, A
87	19.2	76.8	600	9	US-10-972-079-15243	Sequence 15243, A
88	19.2	76.8	600	9	US-10-972-079-15244	Sequence 15244, A
89	19.2	76.8	600	9	US-10-972-079-15245	Sequence 15245, A
90	19.2	76.8	600	9	US-10-972-079-15246	Sequence 15246, A
91	19.2	76.8	600	9	US-10-972-079-15247	Sequence 15247, A
92	19.2	76.8	600	9	US-10-972-079-15248	Sequence 15248, A
93	19.2	76.8	600	9	US-10-972-079-15249	Sequence 15249, A
94	19.2	76.8	600	9	US-10-972-079-15250	Sequence 15250, A
95	19.2	76.8	600	9	US-10-972-079-15251	Sequence 15251, A
96	19.2	76.8	600	9	US-10-972-079-15252	Sequence 15252, A

```
97 18.8 75.2 959 6 US-10-027-632-151970 Sequence 151970,
98 18.8 75.2 3246 9 US-10-887-5538-988 Sequence 988, App
99 18.8 75.2 25619 3 US-09-984-429-576 Sequence 576, App
100 18.8 75.2 25619 3 US-09-908-711-143 Sequence 143, App
101 18.8 75.2 25619 3 US-09-764-898-302 Sequence 302, App
102 18.8 75.2 25619 3 US-09-764-891-6213 Sequence 6213, App
103 18.8 75.2 32523 3 US-09-984-429-296 Sequence 296, App
104 18.8 75.2 45664 8 US-10-741-600-17881 Sequence 17881, A
105 18.8 75.2 51001 6 US-10-189-268-11 Sequence 11, Appl
106 18.8 75.2 51705 7 US-10-052-482-229 Sequence 229, App
107 18.6 74.4 161 7 US-10-424-599-88693 Sequence 88693, A
108 18.6 74.4 232 7 US-10-424-599-65191 Sequence 65191, A
109 18.6 74.4 254 7 US-10-424-599-16082 Sequence 16082, A
110 18.6 74.4 351 7 US-10-424-599-7689 Sequence 7689, App
111 18.6 74.4 350 4 US-09-925-065A-547662 Sequence 547662,
112 18.6 74.4 390 4 US-09-925-065A-547663 Sequence 547663,
113 18.6 74.4 391 8 US-10-425-115-11211 Sequence 11211, A
114 18.6 74.4 402 4 US-09-925-065A-289646 Sequence 289646,
115 18.6 74.4 402 4 US-09-925-065A-658278 Sequence 658278,
116 18.6 74.4 408 4 US-09-925-065A-658276 Sequence 658276,
117 18.6 74.4 408 4 US-09-925-065A-658277 Sequence 658277,
118 18.6 74.4 422 4 US-09-925-065A-639493 Sequence 639493,
119 18.6 74.4 437 4 US-09-925-065A-470467 Sequence 470467,
120 18.6 74.4 437 4 US-09-925-065A-470468 Sequence 470468,
121 18.6 74.4 437 4 US-09-925-065A-470469 Sequence 470469,
122 18.6 74.4 506 4 US-09-925-065A-35506 Sequence 35506, A
123 18.6 74.4 515 4 US-09-925-065A-257656 Sequence 257656,
124 18.6 74.4 519 7 US-10-425-114-14766 Sequence 14766, A
125 18.6 74.4 537 5 US-10-027-632-257543 Sequence 257543,
126 18.6 74.4 537 6 US-10-027-632-257543 Sequence 257543,
127 18.6 74.4 570 4 US-09-925-065A-261160 Sequence 261160,
128 18.6 74.4 570 4 US-09-925-065A-261161 Sequence 261161,
129 18.6 74.4 579 4 US-09-925-065A-364343 Sequence 364343,
130 18.6 74.4 579 4 US-09-925-065A-364345 Sequence 364345,
131 18.6 74.4 597 3 US-09-864-761-8960 Sequence 8960, App
132 18.6 74.4 599 9 US-10-972-079-50639 Sequence 50639, A
133 18.6 74.4 601 4 US-09-925-065A-28010 Sequence 28010, A
134 18.6 74.4 604 4 US-09-925-065A-452703 Sequence 452703,
135 18.6 74.4 604 4 US-09-925-065A-452704 Sequence 452704,
136 18.6 74.4 604 4 US-09-925-065A-452705 Sequence 452705,
137 18.6 74.4 605 4 US-09-925-065A-350420 Sequence 350420,
138 18.6 74.4 605 4 US-09-925-065A-350421 Sequence 350421,
139 18.6 74.4 606 4 US-09-925-065A-277788 Sequence 277788,
140 18.6 74.4 612 7 US-10-767-701-28163 Sequence 28163, A
141 18.6 74.4 616 4 US-09-925-065A-867090 Sequence 867090,
142 18.6 74.4 650 3 US-09-770-149-527 Sequence 527, App
143 18.6 74.4 684 4 US-09-925-065A-909508 Sequence 909508,
144 18.6 74.4 728 7 US-10-424-599-77061 Sequence 77061, A
145 18.6 74.4 761 7 US-10-424-599-142413 Sequence 142413,
146 18.6 74.4 836 5 US-10-027-632-171388 Sequence 171388,
147 18.6 74.4 836 6 US-10-027-632-171388 Sequence 171388,
148 18.6 74.4 860 5 US-10-198-846-4969 Sequence 4969, App
149 18.6 74.4 898 7 US-10-424-599-3082 Sequence 3082, App
150 18.6 74.4 1066 7 US-10-437-963-101667 Sequence 101667,
```

ALIGNMENTS

```
RESULT 1
US-10-681-773-6
; Sequence 6, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 6
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-6
```

```
Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred No. 2.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
Db 1 AAAAAAAAAAGCATGTTGTGACAC 25
```

```
RESULT 2
US-10-891-260-3906
; Sequence 3906, Application US/10891260
; Publication No. US20050227244A1
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc.
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; TITLE OF INVENTION: Methods of Analysis of Human Polymorphisms
; FILE REFERENCE: 3522.3
; CURRENT APPLICATION NUMBER: US/10/891,260
; CURRENT FILING DATE: 2004-07-13
; PRIOR APPLICATION NUMBER: 10/681,773
; PRIOR FILING DATE: 2003-10-07
; NUMBER OF SEQ ID NOS: 10244
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3906
; LENGTH: 33
; TYPE: DNA
; ORGANISM: homo sapien
US-10-891-260-3906
```

```
Query Match 98.4%; Score 24.6; DB 9; Length 33;
Best Local Similarity 96.0%; Pred. No. 3.4;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
Db 1 AAAAAAAAAAGCATGTTGTGACAC 25
```

```
RESULT 3
US-10-027-632-66793
; Sequence 66793, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
```



```
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 66793
;; LENGTH: 515
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-66793
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 515;
Best Local Similarity 96.0%; Pred. No. 5.4;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGATGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGATGTGACAC 32
```

```
RESULT 4
US-10-027-632-66794
```

```
;; Sequence 66794, Application US/10027632
;; Publication No. US20020198371A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; POLYMORPHISMS IN THE HUMAN GENOME
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; PRIOR FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 66794
;; LENGTH: 515
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-66794
```

```
Query Match          98.4%; Score 24.6; DB 5; Length 515;
Best Local Similarity 96.0%; Pred. No. 5.4;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGATGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGATGTGACAC 32
```

```
RESULT 5
US-10-027-632-66793
```

```
;; Sequence 66793, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; POLYMORPHISMS IN THE HUMAN GENOME
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; PRIOR FILING DATE: 2002-04-30
```

```
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 66793
;; LENGTH: 515
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-66793
```

```
Query Match          98.4%; Score 24.6; DB 6; Length 515;
Best Local Similarity 96.0%; Pred. No. 5.4;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGATGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGATGTGACAC 32
```

```
RESULT 6
US-10-027-632-66794
```

```
;; Sequence 66794, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; POLYMORPHISMS IN THE HUMAN GENOME
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; PRIOR FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 66794
;; LENGTH: 515
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-66794
```

```
Query Match          98.4%; Score 24.6; DB 6; Length 515;
Best Local Similarity 96.0%; Pred. No. 5.4;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGCATGATGTGACAC 25
    |||||
Db 8 AAAAAAAAAAGCATGATGTGACAC 32
```

```
RESULT 7
US-10-027-632-34604
; Sequence 34604, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34604
; TYPE: DNA
; LENGTH: 651
; ORGANISM: Human
US-10-027-632-34604

Query Match          98.4%; Score 24.6; DB 5; Length 651;
Best Local Similarity 96.0%; Pred. No. 5.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGCAC 25
DB 169 AAAAAAAAAAGCATGATGTGCAC 193

RESULT 8
US-10-027-632-34605
; Sequence 34605, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34605
; TYPE: DNA
; LENGTH: 651
; ORGANISM: Human
```

```
US-10-027-632-34605

Query Match          98.4%; Score 24.6; DB 5; Length 651;
Best Local Similarity 96.0%; Pred. No. 5.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGCAC 25
DB 169 AAAAAAAAAAGCATGATGTGCAC 193

RESULT 9
US-10-027-632-34604
; Sequence 34604, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34604
; TYPE: DNA
; LENGTH: 651
; ORGANISM: Human
US-10-027-632-34604

Query Match          98.4%; Score 24.6; DB 6; Length 651;
Best Local Similarity 96.0%; Pred. No. 5.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGCAC 25
DB 169 AAAAAAAAAAGCATGATGTGCAC 193

RESULT 10
US-10-027-632-34605
; Sequence 34605, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34605
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34605
```

```

Query Match          98.4%; Score 24.6; DB 6; Length 651;
Best Local Similarity 96.0%; Pred. No. 5.6;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
Db 169 AAAAAAAAAAGCATGATTGTGACAC 193
```

```

RESULT 11
; Sequence 5, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 5
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-5
```

```

Query Match          93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 9.9;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
Db 1 AAAAAAAAAAGCATGATTGTGACAC 25
```

```

RESULT 12
US-10-681-773-42
; Sequence 42, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
```

```

; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 42
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-42
```

```

Query Match          88.0%; Score 22; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 4 AAAAAAAAAAGCATGATTGTGACAC 25
Db 1 AAAAAAAAAAGCATGATTGTGACAC 22
```

```

RESULT 13
US-10-681-773-115
; Sequence 115, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 115
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-115
```

```

Query Match          84.0%; Score 21; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 5 AAAAAAAAAAGCATGATTGTGACAC 25
Db 1 AAAAAAAAAAGCATGATTGTGACAC 21
```

```

RESULT 14
US-09-997-722-160
; Sequence 160, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 160
; LENGTH: 32404
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
/ NAME/KEY: misc.feature
/ LOCATION: (384)..(403)
/ OTHER INFORMATION: "n" at positions 384 through 403 can be any base.
US-09-997-722-160
```

```
Query Match      84.0%; Score 21; DB 3; Length 32404;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTG 21
         |||||
Db       302 AAAAAAAAAAGCATGATTGTG 322
```

```
RESULT 15
US-09-938-842A-1519/c
/ Sequence 1519, Application US/09938842A
/ Patent No. US20020160378A1
/ GENERAL INFORMATION:
/ APPLICANT: Harper, Jeff
/ APPLICANT: Kreps, Joel
/ APPLICANT: Wang, Xun
/ APPLICANT: Zhu, Tong
/ TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
/ FILE REFERENCE: SCRIPT300-3
/ CURRENT APPLICATION NUMBER: US/09/938,842A
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 60/227,866
/ PRIOR FILING DATE: 2000-08-24
/ PRIOR APPLICATION NUMBER: US 60/264,647
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/300,111
/ PRIOR FILING DATE: 2001-06-22
/ NUMBER OF SEQ ID NOS: 5379
/ SEQ ID NO 1519
/ LENGTH: 2013
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-09-938-842A-1519
```

```
Query Match      83.2%; Score 20.8; DB 3; Length 2013;
Best Local Similarity 91.7%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGACA 24
         |||||
Db       1329 AAAAAAAAAAGCATGATTGTGACA 1306
```

```
RESULT 16
US-09-938-842A-1519/c
/ Sequence 1519, Application US/09938842A
/ Publication No. US20040009476A9
/ GENERAL INFORMATION:
/ APPLICANT: Harper, Jeff
/ APPLICANT: Kreps, Joel
/ APPLICANT: Wang, Xun
/ APPLICANT: Zhu, Tong
/ TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
/ FILE REFERENCE: SCRIPT300-3
/ CURRENT APPLICATION NUMBER: US/09/938,842A
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 60/227,866
/ PRIOR FILING DATE: 2000-08-24
/ PRIOR APPLICATION NUMBER: US 60/264,647
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/300,111
/ PRIOR FILING DATE: 2001-06-22
/ NUMBER OF SEQ ID NOS: 5379
/ SEQ ID NO 1519
/ LENGTH: 2013
```

```
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-09-938-842A-1519
```

```
Query Match      83.2%; Score 20.8; DB 3; Length 2013;
Best Local Similarity 91.7%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGACA 24
         |||||
Db       1329 AAAAAAAAAAGCATGATTGTGACA 1306
```

```
RESULT 17
US-10-719-993-6986/c
/ Sequence 6986, Application US/10719993
/ Publication No. US20040265849A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CL001496
/ CURRENT APPLICATION NUMBER: US/10/719,993
/ PRIOR FILING DATE: 2003-11-24
/ NUMBER OF SEQ ID NOS: 55342
/ SOFTWARE: FastSeq for windows Version 4.0
/ SEQ ID NO 6986
/ LENGTH: 130030
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1)..(130030)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6986
```

```
Query Match      83.2%; Score 20.8; DB 8; Length 130030;
Best Local Similarity 91.7%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGACA 24
         |||||
Db       21508 AAAAAAAAAAGCATGATTGTGACA 21485
```

```
RESULT 18
US-10-719-993-6939
/ Sequence 6939, Application US/10719993
/ Publication No. US20040265849A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CL001496
/ CURRENT APPLICATION NUMBER: US/10/719,993
/ PRIOR FILING DATE: 2003-11-24
/ NUMBER OF SEQ ID NOS: 55342
/ SOFTWARE: FastSeq for windows Version 4.0
/ SEQ ID NO 6939
/ LENGTH: 193757
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1)..(193757)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-719-993-6939
```

```
Query Match      83.2%; Score 20.8; DB 8; Length 193757;
Best Local Similarity 91.7%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGACA 24
```

```
Db      85989 AAAAAAAAAAGCATGTTGTGTCA 86012
|||||
RESULT 19
US-10-719-993-6940/c
; Sequence 6940, Application US/10719993
; Publication No. US20040265849a1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6940
; LENGTH: 1790242
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(1790242)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6940

Query Match      83.2%; Score 20.8; DB 8; Length 1790242;
Best Local Similarity 91.7%; Pred. No. 6.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAAGCATGTTGTGACA 24
|||||
Db      705163 AAAAAAAAAAGCATGTTGTGTCA 705140
|||||

RESULT 20
US-10-681-773-41
; Sequence 41, Application US/10681773
; Publication No. US20040146890a1
; GENERAL INFORMATION:
; APPLICANT: Matczuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 41
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-41

Query Match      81.6%; Score 20.4; DB 7; Length 25;
Best Local Similarity 95.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db      1 AAAAAAAAAAGCATGTTGTGACAC 22
|||||

RESULT 21
US-10-425-115-102172
; Sequence 102172, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 102172
; LENGTH: 725
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_24689C.1
US-10-425-115-102172

Query Match      80.8%; Score 20.2; DB 8; Length 725;
Best Local Similarity 88.0%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db      697 AAAAAAAAAAGCACAAATTTTGACAC 721
|||||

RESULT 22
US-10-394-948-19/c
; Sequence 19, Application US/10394948
; Publication No. US20040023267A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: No. US20040023267A1e1 Compositions and Methods in Cancer
; FILE REFERENCE: 529452000900
; CURRENT APPLICATION NUMBER: US/10/394,948
; CURRENT FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: US 60/367,025
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 25032
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(25032)
; OTHER INFORMATION: n = A,T,C or G
US-10-394-948-19

Query Match      80.8%; Score 20.2; DB 7; Length 25032;
Best Local Similarity 88.0%; Pred. No. 5.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAAGCATGTTGTGACAC 25
|||||
Db      17554 AAAAAAAAAATCATTAATGTGACTC 17530
|||||

RESULT 23
US-10-052-482-1/c
; Sequence 1, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
```

```

; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 25032
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (623)..(642)
; OTHER INFORMATION: "n" at positions 613 to 642 can be any base.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1720)..(2727)
; OTHER INFORMATION: "n" at positions 1720 to 2727 can be any base.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9567)..(9586)
; OTHER INFORMATION: "n" at positions 9567 to 9586 can be any base.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13042)..(13295)
; OTHER INFORMATION: "n" at positions 13042 to 13295 can be any base.
; US-10-052-482-1

Query Match      80.8%; Score 20.2; DB 7; Length 25032;
Best Local Similarity 88.0%; Pred. No. 5.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
Db 17554 AAAAAAAAAATCATTTATGTGACTC 17530

RESULT 24
US-10-697-828-11/c
; Sequence 11, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PaateSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 160552
; TYPE: DNA
; ORGANISM: homo sapiens
; US-10-697-828-11

Query Match      80.8%; Score 20.2; DB 8; Length 160552;
Best Local Similarity 88.0%; Pred. No. 8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
Db 140431 AAAAAAAAAAGCCTTATTGAGACAC 140407

RESULT 25
US-10-681-773-316
; Sequence 316, Application US/10681773

; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Wei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methode for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 316
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
; US-10-681-773-316

Query Match      80.0%; Score 20; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AAAAAAGCATGATTGTGCAC 25
Db 1 AAAAAAGCATGATTGTGCAC 20

RESULT 26
US-10-425-115-16636/c
; Sequence 16636, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 16636
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_115170C.1
; US-10-425-115-16636

Query Match      79.2%; Score 19.8; DB 8; Length 279;
Best Local Similarity 91.3%; Pred. No. 4.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 23
Db 222 AAAAAAAAAAGTATGTGAGAC 200

RESULT 27
US-09-925-065A-706877/c
; Sequence 706877, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
```

```
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 706877
; LENGTH: 788
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-706877

Query Match          79.2%; Score 19.8; DB 4; Length 788;
Best Local Similarity 91.3%; Pred. No. 4.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAAGCATGATTGTGACA 24
Db 590 AAAAAATCATGATTGTGTGACA 568

RESULT 28
US-10-484-577-683
; Sequence 683, Application US/10484577
; Publication No. US20050032724A1
; GENERAL INFORMATION:
; APPLICANT: EPIDAUROS Biotechnology Aktiengesellschaft
; TITLE OF INVENTION: Means and methods for improved treatment of cancer based on UGT1A
; FILE REFERENCE: F2285PCT-1
; CURRENT APPLICATION NUMBER: US/10/484,577
; CURRENT FILING DATE: 2004-01-22
; PRIOR APPLICATION NUMBER: PCT/EP 02/08220
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: EP 01 11 7608.8
; PRIOR FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: EP 02011710.7
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 683
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 683
; LENGTH: 177380
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-484-577-683

Query Match          79.2%; Score 19.8; DB 8; Length 177380;
Best Local Similarity 91.3%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGAC 23
Db 96631 AAGAAAGAAAGCATGATTGTGAC 96653

RESULT 29
US-10-087-192-205/c
; Sequence 205, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 205
; LENGTH: 493631
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493631)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-205

Query Match          79.2%; Score 19.8; DB 5; Length 493631;
Best Local Similarity 91.3%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGAC 23
Db 303132 AAAAAAATCATGATTGTGTGAC 303110

RESULT 30
US-10-681-773-113
; Sequence 113, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 113
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-113

Query Match          77.6%; Score 19.4; DB 7; Length 25;
Best Local Similarity 95.2%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 5 AAAAAAGCATGATTGTGACAC 25
Db 1 AAAAAAGCATGACTGTGACAC 21

RESULT 31
US-09-867-701-4792
; Sequence 4792, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglatae, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4792
; LENGTH: 329
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4792

Query Match      77.6%; Score 19.4; DB 3; Length 329;
Best Local Similarity 95.2%; Pred. No. 6.1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTG 21
Db 68 AAAAAAAAAAGCAAGATTGTG 88

RESULT 32
US-09-962-436-507
; Sequence 507, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppel, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 507
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-507

Query Match      77.6%; Score 19.4; DB 3; Length 356;
Best Local Similarity 95.2%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTG 21
Db 68 AAAAAAAAAAGCAAGATTGTG 88

RESULT 33
US-09-867-701-3652
; Sequence 3652, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3652
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-3652

Query Match      77.6%; Score 19.4; DB 3; Length 356;
Best Local Similarity 95.2%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 AAAAAAAAAAGCATGTTGTG 21
Db 68 AAAAAAAAAAGCAAGATTGTG 88

RESULT 34
US-10-843-641A-2966
; Sequence 2966, Application US/10843641A
; Publication No. US20050064454A1
; GENERAL INFORMATION:
; APPLICANT: Avalon Pharmaceuticals, Inc.
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; FILE REFERENCE: 689290-189
; CURRENT APPLICATION NUMBER: US/10/843,641A
; CURRENT FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: US/09/873,367
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,436
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,832
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/964,824
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/09/967,768
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/09/968,007
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,347
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,708
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 8447
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2966
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-843-641A-2966

Query Match      77.6%; Score 19.4; DB 9; Length 356;
Best Local Similarity 95.2%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTG 21
Db 68 AAAAAAAAAAGCAAGATTGTG 88

RESULT 35
US-09-867-701-10644
; Sequence 10644, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10644
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Homo sapien
```



```
FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(377)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-10644

Query Match      77.6%; Score 19.4; DB 3; Length 377;
Best Local Similarity 95.2%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTG 21
Db 68 AAAAAAAAAAGCAAGATTGTG 88

RESULT 36
US-09-867-701-4991/C
; Sequence 4991, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agilate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4991
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4991

Query Match      77.6%; Score 19.4; DB 3; Length 393;
Best Local Similarity 95.2%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTG 21
Db 342 AAAAAAAAAAGCAAGATTGTG 322

RESULT 37
US-09-867-701-10729/C
; Sequence 10729, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agilate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10729
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-10729

Query Match      77.6%; Score 19.4; DB 3; Length 396;
Best Local Similarity 95.2%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTG 21
Db 343 AAAAAAAAAAGCAAGATTGTG 323

RESULT 38
US-10-956-157-2976
; Sequence 2976, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounes, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2976
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-2976

Query Match      77.6%; Score 19.4; DB 9; Length 643;
Best Local Similarity 95.2%; Pred. No. 6.8e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTG 21
Db 93 AAAAAAAAAAGCAAGATTGTG 113

RESULT 39
US-10-956-157-8211/C
; Sequence 8211, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounes, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8211
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-8211

Query Match      77.6%; Score 19.4; DB 9; Length 643;
Best Local Similarity 95.2%; Pred. No. 6.8e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTG 21
Db 551 AAAAAAAAAAGCAAGATTGTG 531

RESULT 40
US-10-175-523-143
; Sequence 143, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: laeng, Pascal
; APPLICANT: Paleyman, Michael
; APPLICANT: Rajan, Pithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
```

```
/ FILE REFERENCE: 3235/10795-US3
/ CURRENT APPLICATION NUMBER: US/10/175,523
/ CURRENT FILING DATE: 2002-06-18
/ PRIOR APPLICATION NUMBER: US 60/299,151
/ PRIOR FILING DATE: 2001-06-18
/ PRIOR APPLICATION NUMBER: US 60/317,828
/ PRIOR FILING DATE: 2001-09-07
/ PRIOR APPLICATION NUMBER: US 60/325,150
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: US 60/333,047
/ PRIOR FILING DATE: 2001-11-14
/ PRIOR APPLICATION NUMBER: US 60/349,936
/ PRIOR FILING DATE: 2002-01-18
/ PRIOR APPLICATION NUMBER: US 60/361,834
/ NUMBER OF SEQ ID NOS: 197
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 143
/ LENGTH: 3434
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-175-523-143

Query Match          77.6%; Score 19.4; DB 5; Length 3434;
Best Local Similarity 95.2%; Pred. No. 8.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 3407 AAAAAAAAAAGCCTGATTGTG 3427

RESULT 41
US-11-099-266-143
/ Sequence 143, Application US/11099266
/ Publication No. US2005018143A1
/ GENERAL INFORMATION:
/ APPLICANT: Brockman, Jeffrey
/ APPLICANT: Evans, David
/ APPLICANT: Hook, Derek
/ APPLICANT: Klimczak, Leszek
/ APPLICANT: Laeng, Pascal
/ APPLICANT: Palfreyman, Michael
/ APPLICANT: Rajan, Pithi
/ TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
/ FILE REFERENCE: 03235/1003795-US4
/ CURRENT APPLICATION NUMBER: US/11/099,266
/ CURRENT FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: US 10/175,523
/ PRIOR FILING DATE: 2002-06-18
/ PRIOR APPLICATION NUMBER: US 60/299,151
/ PRIOR FILING DATE: 2001-06-18
/ PRIOR APPLICATION NUMBER: US 60/317,828
/ PRIOR FILING DATE: 2001-09-07
/ PRIOR APPLICATION NUMBER: US 60/325,150
/ PRIOR FILING DATE: 2001-09-25
/ PRIOR APPLICATION NUMBER: US 60/333,047
/ PRIOR FILING DATE: 2001-11-14
/ PRIOR APPLICATION NUMBER: US 60/349,936
/ PRIOR FILING DATE: 2002-01-18
/ PRIOR APPLICATION NUMBER: US 60/361,834
/ PRIOR FILING DATE: 2002-03-04
/ NUMBER OF SEQ ID NOS: 197
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 143
/ LENGTH: 3434
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-099-266-143

Query Match          77.6%; Score 19.4; DB 10; Length 3434;
Best Local Similarity 95.2%; Pred. No. 8.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 3407 AAAAAAAAAAGCCTGATTGTG 3427

RESULT 42
US-10-322-281-45/c
/ Sequence 45, Application US/10322281
/ Publication No. US20040126762A1
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc S. Malandro
/ TITLE OF INVENTION: Novel Compositions and Methods in Cancer
/ FILE REFERENCE: 52945200100
/ CURRENT APPLICATION NUMBER: US/10/322,281
/ CURRENT FILING DATE: 2002-12-17
/ NUMBER OF SEQ ID NOS: 866
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 45
/ LENGTH: 219352
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(219352)
/ OTHER INFORMATION: n = A,T,C or G
US-10-322-281-45

Query Match          77.6%; Score 19.4; DB 7; Length 219352;
Best Local Similarity 95.2%; Pred. No. 1.7e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 165651 AAAAAAAAAATGCATGATTGTG 165631

RESULT 43
US-10-021-323-3648/c
/ Sequence 3648, Application US/10021323
/ Publication No. US20040123340A1
/ GENERAL INFORMATION:
/ APPLICANT: Deikman, Jill
/ APPLICANT: Feng, Paul C.C.
/ APPLICANT: Fincher, Karen L.
/ APPLICANT: Ziegler, Todd E.
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants
/ FILE REFERENCE: 38-21(52274)B
/ CURRENT APPLICATION NUMBER: US/10/021,323
/ CURRENT FILING DATE: 2001-12-12
/ PRIOR APPLICATION NUMBER: US 60/255, 619
/ PRIOR FILING DATE: 2000-12-14
/ NUMBER OF SEQ ID NOS: 17880
/ SEQ ID NO 3648
/ LENGTH: 163
/ TYPE: DNA
/ ORGANISM: Gossypium hirsutum
/ FEATURE:
/ OTHER INFORMATION: Clone ID: LIB3825-028-06-K6-A2
US-10-021-323-3648

Query Match          76.8%; Score 19.2; DB 7; Length 163;
Best Local Similarity 87.5%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 130 AAAAAAAAAACAGCATGATTAAAGACA 107

RESULT 44
```

```
US-10-021-323-3727
; Sequence 3727, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021.323
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 3727
; LENGTH: 166
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB8625-028-Q6-N6-A2
US-10-021-323-3727

Query Match          76.8%; Score 19.2; DB 7; Length 166;
Best Local Similarity 87.5%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 32 AAAAAAAAAACGACATTTAGACA 55

RESULT 45
US-10-027-632-300127
; Sequence 300127, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218, 006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198, 676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193, 483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185, 218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167, 363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156, 358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146, 002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300127
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-300127

Query Match          76.8%; Score 19.2; DB 5; Length 412;
Best Local Similarity 87.5%; Pred. No. 7.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 6 AAAAAAAAAAGATTTGTGACA 29
```

```
RESULT 46
US-10-027-632-300127
; Sequence 300127, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218, 006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198, 676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193, 483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185, 218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167, 363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156, 358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146, 002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300127
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-300127

Query Match          76.8%; Score 19.2; DB 6; Length 412;
Best Local Similarity 87.5%; Pred. No. 7.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 6 AAAAAAAAAAGATTTGTGACA 29

RESULT 47
US-09-764-891-716
; Sequence 716, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 716
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (45)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (72)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-716
```

```
Query Match      76.8%; Score 19.2; DB 3; Length 414;
Best Local Similarity 87.5%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGCATGTTGTGACA 24
        |||||
        6 AAAAAAAAAAATCATGATTGTGAAA 158

Db      135 AAAAAAAAAAATCATGATTGTGAAA 158

RESULT 48
US-10-027-632-62999
; Sequence 62999, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62999
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-62999

Query Match      76.8%; Score 19.2; DB 5; Length 438;
Best Local Similarity 87.5%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGCATGTTGTGACA 24
        |||||
        6 AAAAAAAAAAATCATGTTGTGACA 29

Db      6 AAAAAAAAAAATCATGTTGTGACA 29

RESULT 49
US-10-027-632-62999
; Sequence 62999, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4453
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-62999

Query Match      76.8%; Score 19.2; DB 6; Length 438;
Best Local Similarity 87.5%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGCATGTTGTGACA 24
        |||||
        6 AAAAAAAAAAATCATGTTGTGACA 29

Db      6 AAAAAAAAAAATCATGTTGTGACA 29

RESULT 50
US-09-864-761-4453
; Sequence 4453, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4453
```

```

; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007859.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.64
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.62
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.86
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92
US-09-864-761-4453

```

```

Query Match      76.8%; Score 19.2; DB 3; Length 464;
Best Local Similarity 87.5%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAGCATGATTGACAC 24
        |||||
Db      137 AAAAAAAAAACCATGATTGAAA 160

```

Search completed: December 14, 2005, 08:46:33
 Job time : 379.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 ; Search time 373.2 Seconds

(without alignments)
553.951 Million cell updates/sec

Title: US-10-681-773-7

Perfect score: 25
Sequence: 1 aaaaaaaaaaagtcacatcagata 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications NA Main:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*

6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	7	US-10-681-773-7
2	25	100.0	418	4	US-09-925-065A-169175
3	24.6	98.4	33	9	US-10-891-260-5923
4	24.6	98.4	425	4	US-09-925-065A-169174
5	24	96.0	25	7	US-10-681-773-14
6	24	96.0	25	7	US-10-681-773-62158
7	23.4	93.6	25	7	US-10-681-773-8
8	23	92.0	25	7	US-10-681-773-18418
9	22.4	89.6	25	7	US-10-681-773-15
10	22.4	89.6	25	7	US-10-681-773-62159
11	22.4	89.6	524	4	US-09-925-065A-793833
12	21.8	87.2	5933	7	US-10-257-166-59
13	21.8	87.2	10151	6	US-10-311-455-2406
14	21.4	85.6	25	7	US-10-681-773-18419
15	21	84.0	25	7	US-10-681-773-131
16	20.2	80.8	470	4	US-09-925-065A-701169
17	20.2	80.8	515	4	US-09-925-065A-610806
18	20.2	80.8	595	4	US-09-925-065A-307635
19	20.2	80.8	600	9	US-10-972-079-84549
20	20.2	80.8	600	9	US-10-972-079-84550
21	20.2	80.8	1043	4	US-09-925-065A-671634
22	20.2	80.8	1270	4	US-09-925-065A-691066
23	20.2	80.8	12542	3	US-09-764-864-1774

24	20.2	80.8	13558	9	US-10-257-763-5	Sequence 5, Appli
25	20.2	80.8	17893	6	US-10-311-455-1337	Sequence 1337, Ap
26	20.2	80.8	109586	5	US-10-087-192-220	Sequence 220, App
27	19.8	79.2	348	3	US-09-954-456-1729	Sequence 1729, Ap
28	19.8	79.2	348	9	US-10-843-61A-4756	Sequence 4756, Ap
29	19.8	79.2	483	5	US-10-027-632-324607	Sequence 324607,
30	19.8	79.2	483	6	US-10-027-632-324607	Sequence 324607,
31	19.8	79.2	495	5	US-10-027-632-82159	Sequence 82159, A
32	19.8	79.2	495	5	US-10-027-632-302138	Sequence 302138,
33	19.8	79.2	495	6	US-10-027-632-82159	Sequence 82159, A
34	19.8	79.2	495	6	US-10-027-632-302138	Sequence 302138,
35	19.8	79.2	512	4	US-09-925-065A-338896	Sequence 338896,
36	19.8	79.2	512	4	US-09-925-065A-338896	Sequence 338896,
37	19.8	79.2	566	4	US-09-925-065A-389484	Sequence 389484,
38	19.8	79.2	668	4	US-09-925-065A-375699	Sequence 375699,
39	19.8	79.2	1662	4	US-09-925-065A-510209	Sequence 510209,
40	19.8	79.2	1662	4	US-09-925-065A-51386	Sequence 51386, A
41	19.8	79.2	1662	4	US-09-925-065A-51387	Sequence 51387, A
42	19.8	79.2	6974	8	US-10-723-860-2033	Sequence 2033, Ap
43	19.8	79.2	7032	3	US-09-974-298-124	Sequence 124, App
44	19.8	79.2	7835	8	US-10-723-860-6323	Sequence 6323, Ap
45	19.8	79.2	10514	3	US-09-764-877-3470	Sequence 3470, Ap
46	19.8	79.2	10514	6	US-10-242-515-3470	Sequence 3470, Ap
47	19.4	77.6	325	7	US-10-681-773-134	Sequence 134, App
48	19.4	77.6	532	4	US-09-925-065A-574492	Sequence 574492,
49	19.4	77.6	532	4	US-09-925-065A-574493	Sequence 574493,
50	19.4	77.6	559	4	US-09-925-065A-394568	Sequence 394568,
51	19.4	77.6	559	4	US-09-925-065A-394569	Sequence 394569,
52	19.4	77.6	991	8	US-10-739-830-3455	Sequence 810110,
53	19.4	77.6	991	8	US-10-739-830-3455	Sequence 3453, Ap
54	19.4	77.6	51001	6	US-10-189-268-11	Sequence 11, Appli
55	19.4	77.6	235033	5	US-10-301-844-1	Sequence 1, Appli
56	19.4	77.6	235033	5	US-10-301-844-1	Sequence 2, Appli
57	19.2	76.8	334	7	US-10-425-115-2021	Sequence 2021, Ap
58	19.2	76.8	336	7	US-10-425-115-2021	Sequence 134659,
59	19.2	76.8	442	8	US-10-425-115-153756	Sequence 133576,
60	19.2	76.8	442	4	US-09-925-065A-495039	Sequence 495039,
61	19.2	76.8	535	4	US-09-925-065A-425519	Sequence 425519,
62	19.2	76.8	535	4	US-09-925-065A-425520	Sequence 425520,
63	19.2	76.8	535	4	US-09-925-065A-425521	Sequence 425521,
64	19.2	76.8	546	4	US-09-925-065A-544535	Sequence 544535,
65	19.2	76.8	577	5	US-10-027-632-92882	Sequence 92882, A
66	19.2	76.8	577	5	US-10-027-632-303295	Sequence 303295,
67	19.2	76.8	577	6	US-10-027-632-92882	Sequence 92882, A
68	19.2	76.8	584	7	US-10-027-632-303295	Sequence 303295,
69	19.2	76.8	584	8	US-10-425-115-33678	Sequence 33678, A
70	19.2	76.8	600	4	US-09-925-065A-537472	Sequence 537472,
71	19.2	76.8	601	4	US-09-925-065A-333717	Sequence 333717,
72	19.2	76.8	611	5	US-10-027-632-186490	Sequence 186490,
73	19.2	76.8	611	5	US-10-027-632-186491	Sequence 186491,
74	19.2	76.8	611	5	US-10-027-632-186491	Sequence 186491,
75	19.2	76.8	611	5	US-10-027-632-186491	Sequence 186491,
76	19.2	76.8	611	6	US-10-027-632-186491	Sequence 186491,
77	19.2	76.8	611	6	US-10-027-632-186491	Sequence 186491,
78	19.2	76.8	2567	7	US-10-425-114-29775	Sequence 22775, A
79	19.2	76.8	3333	7	US-10-424-599-109133	Sequence 109133,
80	19.2	76.8	3333	5	US-10-027-632-118856	Sequence 118856,
81	19.2	76.8	3333	5	US-10-027-632-118857	Sequence 118857,
82	19.2	76.8	3333	6	US-10-027-632-118856	Sequence 118856,
83	19.2	76.8	3333	6	US-10-027-632-118857	Sequence 118857,
84	19.2	76.8	4411	6	US-10-239-642-31	Sequence 31, Appli
85	19.2	76.8	5688	3	US-09-764-869-1740	Sequence 1740, Ap
86	19.2	76.8	5688	3	US-09-764-869-1740	Sequence 5912, Ap
87	19.2	76.8	5688	5	US-10-091-504-1740	Sequence 1740, Ap
88	19.2	76.8	5688	6	US-10-027-577-1740	Sequence 1740, Ap
89	19.2	76.8	31221	10	US-11-097-143-37261	Sequence 37261, A
90	19.2	76.8	65454	7	US-10-293-864-11	Sequence 11, Appli
91	19.2	76.8	121410	8	US-10-741-600-17733	Sequence 17733, A
92	19	76.0	584	4	US-09-925-065A-394567	Sequence 394567,
93	19	76.0	6764	7	US-10-316-242-11	Sequence 11, Appli
94	18.8	75.2	277	7	US-10-424-599-11964	Sequence 11964,
95	18.8	75.2	353	7	US-10-424-599-899	Sequence 899, App
96	18.8	75.2	432	8	US-10-425-115-156620	Sequence 156620,

```
C 97 18.8 75.2 507 7 US-10-778-617-18 Sequence 19, Appl
98 18.8 75.2 580 4 US-09-925-065A-660788 Sequence 60788,
C 99 18.8 75.2 600 4 US-09-925-065A-787713 Sequence 787713,
C 100 18.8 75.2 600 4 US-09-925-065A-787714 Sequence 787714,
C 101 18.8 75.2 608 4 US-09-925-065A-258215 Sequence 258215,
C 102 18.8 75.2 7421 6 US-10-310-154-110 Sequence 110, Appl
C 103 18.8 75.2 7421 9 US-10-732-923-14 Sequence 14, Appl
C 104 18.8 75.2 34389 5 US-10-087-192-1375 Sequence 1375, A
C 105 18.8 75.2 59110 8 US-10-741-600-17742 Sequence 17742, A
C 106 18.8 75.2 193357 6 US-10-085-117-142 Sequence 142, App
C 107 18.6 74.4 201 8 US-10-719-993-20539 Sequence 20539, A
C 108 18.6 74.4 201 8 US-10-719-993-20542 Sequence 20542, A
C 109 18.6 74.4 201 8 US-10-719-993-20544 Sequence 20544, A
C 110 18.6 74.4 201 8 US-10-719-993-20546 Sequence 20546, A
C 111 18.6 74.4 201 8 US-10-719-993-20546 Sequence 20546, A
C 112 18.6 74.4 201 8 US-10-741-600-40373 Sequence 40373, A
C 113 18.6 74.4 201 8 US-10-741-600-40376 Sequence 40376, A
C 114 18.6 74.4 201 8 US-10-741-600-40378 Sequence 40378, A
C 115 18.6 74.4 201 8 US-10-741-600-40380 Sequence 40380, A
C 116 18.6 74.4 201 8 US-10-741-600-40382 Sequence 40382, A
C 117 18.6 74.4 201 8 US-10-741-600-58025 Sequence 58025, A
C 118 18.6 74.4 201 8 US-10-741-600-58035 Sequence 58035, A
C 119 18.6 74.4 201 8 US-10-741-600-58044 Sequence 58044, A
C 120 18.6 74.4 243 7 US-10-424-599-15456 Sequence 15456, A
C 121 18.6 74.4 248 7 US-10-424-599-75388 Sequence 75388, A
C 122 18.6 74.4 265 8 US-10-425-115-55912 Sequence 55912, A
C 123 18.6 74.4 267 8 US-10-425-115-5802 Sequence 5802, Ap
C 124 18.6 74.4 269 8 US-10-425-115-7279 Sequence 7279, Ap
C 125 18.6 74.4 366 7 US-10-424-599-102609 Sequence 102609,
C 126 18.6 74.4 398 8 US-10-425-115-42039 Sequence 42039, A
C 127 18.6 74.4 406 9 US-10-779-943-12294 Sequence 12294, A
C 128 18.6 74.4 431 3 US-09-814-353-21670 Sequence 21670, A
C 129 18.6 74.4 435 7 US-10-191-803-408 Sequence 408, App
C 130 18.6 74.4 441 8 US-10-357-930-17713 Sequence 17713, A
C 131 18.6 74.4 447 4 US-09-925-065A-96822 Sequence 96822, A
C 132 18.6 74.4 448 4 US-09-925-065A-590019 Sequence 590019,
C 133 18.6 74.4 450 5 US-10-027-632-292613 Sequence 292613,
C 134 18.6 74.4 450 6 US-10-027-632-292613 Sequence 292613,
C 135 18.6 74.4 457 4 US-09-925-065A-738216 Sequence 738216,
C 136 18.6 74.4 466 4 US-09-925-065A-657533 Sequence 657533,
C 137 18.6 74.4 470 4 US-09-925-065A-947603 Sequence 947603,
C 138 18.6 74.4 472 4 US-09-925-065A-656466 Sequence 656466,
C 139 18.6 74.4 479 4 US-09-925-065A-172784 Sequence 172784,
C 140 18.6 74.4 484 4 US-09-925-065A-172784 Sequence 172784,
C 141 18.6 74.4 494 4 US-10-027-632-83003 Sequence 83003, A
C 142 18.6 74.4 505 5 US-10-027-632-83003 Sequence 83003, A
C 143 18.6 74.4 507 5 US-10-027-632-229700 Sequence 229700,
C 144 18.6 74.4 507 6 US-10-027-632-229700 Sequence 229700,
C 145 18.6 74.4 510 4 US-09-925-065A-629394 Sequence 629394,
C 146 18.6 74.4 510 4 US-09-925-065A-629394 Sequence 629394,
C 147 18.6 74.4 510 6 US-10-027-632-35279 Sequence 35279, A
C 148 18.6 74.4 514 3 US-10-027-632-35279 Sequence 35279, A
C 149 18.6 74.4 521 3 US-09-814-353-18218 Sequence 18218, A
C 150 18.6 74.4 521 9 US-10-972-079-42267 Sequence 42267, A
```

ALIGNMENTS

```
RESULT 1
US-10-681-773-7
; Sequence 7, Application US/10681773
; Publication No. US2004016890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; TITLE OF INVENTION: Methode for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 7
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-7
```

```
Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 1 AAAAAAAAAAGTCCCAATTCAGATA 25
```

```
RESULT 2
US-09-925-065A-169175
; Sequence 169175, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: PaetSeq for Windows Version 4.0
; SEQ ID NO 169175
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-169175

Query Match 100.0%; Score 25; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 3
US-10-891-260-5923
; Sequence 5923, Application US/10891260
; Publication No. US20050227244A1
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc.
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; TITLE OF INVENTION: Methode of Analysis of Human Polymorphisms
; FILE REFERENCE: 3522.3
; CURRENT APPLICATION NUMBER: US/10/891,260
; CURRENT FILING DATE: 2004-07-13
; PRIOR APPLICATION NUMBER: 10/681,773
; PRIOR FILING DATE: 2003-10-07
; NUMBER OF SEQ ID NOS: 10244
```



```
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 5923
; LENGTH: 33
; TYPE: DNA
; ORGANISM: homo sapien
US-10-681-260-5923
```

```
Query Match          98.4%; Score 24.6; DB 9; Length 33;
Best Local Similarity 96.0%; Pred. No. 4.9;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 5 AAAAAAAAAAGTCCCAATTCAGATA 29
```

RESULT 4

```
US-09-925-065A-169174
; Sequence 169174, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169174
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-169174
```

```
Query Match          98.4%; Score 24.6; DB 4; Length 425;
Best Local Similarity 96.0%; Pred. No. 6.8;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 120 AAAAAAAAAAGTCCCAATTCAGATA 144
```

```
RESULT 5
US-10-681-773-14
; Sequence 14, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```

```
; SEQ ID NO 14
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-14
```

```
Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 1 AAAAAAAAAAGTCCCAATTCAGATA 24
```

RESULT 6

```
US-10-681-773-62158
; Sequence 62158, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 62158
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-62158
```

```
Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 2 AAAAAAAAAAGTCCCAATTCAGAT 25
```

```
RESULT 7
US-10-681-773-8
; Sequence 8, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 8
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-8
```

Query Match 93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
DB 1 AAAAAAAAAAGTCCCAATTCAGATA 25

RESULT 8
US-10-681-773-18418
Sequence 18418, Application US/10681773
Publication No. US20040146890A1

GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
PRIOR FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 18418
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-18418

Query Match 92.0%; Score 23; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
DB 3 AAAAAAAAAAGTCCCAATTCAGA 25

RESULT 9
US-10-681-773-15
Sequence 15, Application US/10681773
Publication No. US20040146890A1

GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
PRIOR FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 15
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-15

Query Match 89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 34;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
DB 1 AAAAAAAAAAGTCCCAATTCAGATA 24

RESULT 10
US-10-681-773-62159
Sequence 62159, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
PRIOR FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 62159
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-62159

Query Match 89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 34;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
DB 2 AAAAAAAAAAGTCCCAATTCAGAT 25

RESULT 11
US-09-925-065A-793833/C
Sequence 793833, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: PaetSBO for Windows Version 4.0
SEQ ID NO 793833
LENGTH: 524
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-793833

Query Match 89.6%; Score 22.4; DB 4; Length 524;
Best Local Similarity 95.8%; Pred. No. 51;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24

Db 224 AAAAAAAAAAGTCCCAATTAGAT 201

RESULT 12
US-10-257-166-59/c
Sequence 59, Application US/10257166
Publication No. US2004002330A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
FILE REFERENCE: 5013.1011
CURRENT APPLICATION NUMBER: US/10/257.166
CURRENT FILING DATE: 2002-10-07
PRIOR APPLICATION NUMBER: PCT/EP01/07470
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-06-29
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 178
SEQ ID NO 58
LENGTH: 5933
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
US-10-257-166-59
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

Query Match 87.2%; Score 21.8; DB 7; Length 5933;
Best Local Similarity 92.0%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAGTA 25
Db 2331 AAAAAAAAAATCCCAATTCAGATA 2307

RESULT 13
US-10-311-455-2406/c
Sequence 2406, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 2406
LENGTH: 10151
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: 9495, 9552
OTHER INFORMATION: n is a or g or c or t
US-10-311-455-2406

Query Match 87.2%; Score 21.8; DB 6; Length 10151;
Best Local Similarity 92.0%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 2331 AAAAAAAAAATCCCAATTCAGATA 2307

RESULT 14
US-10-681-773-18419
Sequence 18419, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681.773
CURRENT FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470.475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417.190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 18419
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-18419

Query Match 85.6%; Score 21.4; DB 7; Length 25;
Best Local Similarity 95.7%; Pred. No. 85;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 3 AAAAAAAAAAGTCCCAATTCAGA 25

RESULT 15
US-10-681-773-131
Sequence 131, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681.773
CURRENT FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470.475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417.190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 131
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-131

Query Match 84.0%; Score 21; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 1 |||||
1 AAAAAAGTCCCAATTCAGATA 21

RESULT 16

US-09-925-065A-701169/C
; Sequence 701169, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 701169
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-701169

Query Match 80.8%; Score 20.2; DB 4; Length 470;
Best Local Similarity 88.0%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGTCCCAATTCAGATA 25
Db 445 AAAAAAAGACCTAATTCAGTTA 421

RESULT 17

US-09-925-065A-610806
; Sequence 610806, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 610806
; LENGTH: 515
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-610806

Query Match 80.8%; Score 20.2; DB 4; Length 515;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGTCCCAATTCAGATA 25
Db 476 AAAAAAAGCCTCAATTCAGATA 500

RESULT 18

US-09-925-065A-307635
; Sequence 307635, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 307635
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-307635

Query Match 80.8%; Score 20.2; DB 4; Length 595;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAGTCCCAATTCAGATA 25
Db 249 AAAAAAAGATCCCAATTCAGTA 273

RESULT 19

US-10-972-079-84549
; Sequence 84549, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFIELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; PRIOR FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 84549
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894382241_1
US-10-972-079-84549

Query Match 80.8%; Score 20.2; DB 9; Length 600;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 462 AAAAAAAAAAGTCCCAATTCAGATA 486

RESULT 20

US-10-972-079-84550
; Sequence 84550, Application US/10972079
; Publication No. US2005015317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 84550
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894382241_2
US-10-972-079-84550

Query Match 80.8%; Score 20.2; DB 9; Length 600;

Best Local Similarity 88.0%; Pred. No. 3.9e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 370 AAAAAAAAAAGTCCCAATTCAGATA 394

RESULT 21

US-09-925-065A-671634
; Sequence 671634, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 671634
; LENGTH: 1043
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-671634

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 973 AAAAAAAAAAGTCCCAATTCAGATA 997

RESULT 22

US-09-925-065A-691066/C
; Sequence 691066, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 691066
; LENGTH: 1270
; TYPE: DNA
; ORGANISM: Homo sapiens

Query Match 80.8%; Score 20.2; DB 4; Length 1270;

Best Local Similarity 88.0%; Pred. No. 4.2e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 1146 AAAAAAAAAAGTCCCAATTCAGATA 1122

RESULT 23

US-09-764-864-1774
; Sequence 1774, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17

```
/ Prior application data removed - consult PALM or file wrapper
/ NUMBER OF SEQ ID NOS: 1792
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 1774
/ LENGTH: 12542
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-764-864-1774

Query Match      80.8%; Score 20.2; DB 3; Length 12542;
Best Local Similarity 88.0%; Pred. No. 5,7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 4239 AAAAAAAAAAGTGCATTTTCAGATA 4263

RESULT 24
US-10-257-763-5/c
/ Sequence 5, Application US/10257763
/ Publication No. US20050147967A1
/ GENERAL INFORMATION:
/ APPLICANT: GOTT, Bernhard et al
/ TITLE OF INVENTION: NOVEL NEURONALLY EXPRESSED PROTEIN AND USE THEREOF
/ FILE REFERENCE: 4266-0101P
/ CURRENT APPLICATION NUMBER: US/10/257,763
/ NUMBER OF SEQ ID NOS: 31
/ SOFTWARE: Patentin version 3.2
/ SEQ ID NO 5
/ LENGTH: 13558
/ TYPE: DNA
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: Unsure
/ LOCATION: (4501)..(4501)
/ OTHER INFORMATION: n=a,c,g,t any unknown or other
US-10-257-763-5

Query Match      80.8%; Score 20.2; DB 9; Length 13558;
Best Local Similarity 88.0%; Pred. No. 5,8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 817 AAAAAAAAAAGTCCCAATTCAGATA 793

RESULT 25
US-10-311-455-1337/c
/ Sequence 1337, Application US/10311455
/ Publication No. US20030143606A1
/ GENERAL INFORMATION:
/ APPLICANT: OLER, Alexander
/ APPLICANT: PIEPERBROCK, Christian
/ APPLICANT: BERLIN, Kurt
/ TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
/ TITLE OF INVENTION: Cytosine methylation
/ FILE REFERENCE: 5013,1014
/ CURRENT APPLICATION NUMBER: US/10/311,455
/ CURRENT FILING DATE: 2002-12-16
/ PRIOR APPLICATION NUMBER: PCT/EP01/07537
/ PRIOR FILING DATE: 2001-07-02
/ PRIOR APPLICATION NUMBER: DE 10032529,7
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: DE 10043826,1
/ PRIOR FILING DATE: 2000-09-01
/ NUMBER OF SEQ ID NOS: 2424
/ SEQ ID NO 1337
/ LENGTH: 17893
/ TYPE: DNA
/ ORGANISM: Artificial Sequence

/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1337

Query Match      80.8%; Score 20.2; DB 6; Length 17893;
Best Local Similarity 88.0%; Pred. No. 6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 8348 AAAAAAAAAATCCCAATTCAGATA 8324

RESULT 26
US-10-087-192-220/c
/ Sequence 220, Application US/10087192
/ Publication No. US20020182586A1
/ GENERAL INFORMATION:
/ APPLICANT: Morris, David W.
/ APPLICANT: Engelhard, Eric K.
/ TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
/ TITLE OF INVENTION: CANCER
/ FILE REFERENCE: 529452000122
/ CURRENT APPLICATION NUMBER: US/10/087,192
/ CURRENT FILING DATE: 2002-03-01
/ PRIOR APPLICATION NUMBER: US 09/747,377
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: US 09/798,586
/ PRIOR FILING DATE: 2001-03-02
/ NUMBER OF SEQ ID NOS: 2059
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 220
/ LENGTH: 109586
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(109586)
/ OTHER INFORMATION: n = A,T,C or G
US-10-087-192-220

Query Match      80.8%; Score 20.2; DB 5; Length 109586;
Best Local Similarity 88.0%; Pred. No. 7,6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 97181 AAAAAAAAAAGTCCCAATTCAGATA 97157

RESULT 27
US-09-954-456-1729/c
/ Sequence 1729, Application US/09954456
/ Patent No. US20020115057A1
/ GENERAL INFORMATION:
/ APPLICANT: Young, Paul
/ TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
/ TITLE OF INVENTION: Sets
/ FILE REFERENCE: 689290-76
/ CURRENT APPLICATION NUMBER: US/09/954,456
/ CURRENT FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: US/60/233,617
/ PRIOR FILING DATE: 2000-09-18
/ PRIOR APPLICATION NUMBER: US/60/234,052
/ PRIOR FILING DATE: 2000-09-20
/ PRIOR APPLICATION NUMBER: US/60/234,923
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US/60/235,134
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US/60/235,637
/ PRIOR FILING DATE: 2000-09-26
/ PRIOR APPLICATION NUMBER: US/60/235,638
/ PRIOR FILING DATE: 2000-09-26
```

```
;; PRIOR APPLICATION NUMBER: US/60/235,711
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: US/60/235,720
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: US/60/235,840
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: US/60/235,863
;; PRIOR FILING DATE: 2000-09-27
;; NUMBER OF SEQ ID NOS: 2276
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 1729
;; LENGTH: 348
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; OTHER INFORMATION: n=a,t,g or c
US-09-954-456-1729
```

```
Query Match          79.2%; Score 19.8; DB 3; Length 348;
Best Local Similarity 91.3%; Pred. No. 5.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGTCCCAATTGCA 23
Db      103 AAAAAATTAAGCCCAATTGCA 81
```

RESULT 28

```
US-10-843-641A-4756/C
;; Sequence 4756, Application US/10843641A
;; Publication No. US20050064454A1
;; GENERAL INFORMATION:
;; APPLICANT: Avalon Pharmaceuticals, Inc.
;; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
;; TITLE OF INVENTION: Signature Gene Sets
;; FILE REFERENCE: 689290-189
;; CURRENT APPLICATION NUMBER: US/10/843,641A
;; CURRENT FILING DATE: 2004-05-12
;; PRIOR APPLICATION NUMBER: US/09/873,367
;; PRIOR FILING DATE: 2001-06-05
;; PRIOR APPLICATION NUMBER: US/09/954,531
;; PRIOR FILING DATE: 2001-09-18
;; PRIOR APPLICATION NUMBER: US/09/954,456
;; PRIOR FILING DATE: 2001-09-25
;; PRIOR APPLICATION NUMBER: US/09/962,436
;; PRIOR FILING DATE: 2001-09-25
;; PRIOR APPLICATION NUMBER: US/09/962,832
;; PRIOR FILING DATE: 2001-09-25
;; PRIOR APPLICATION NUMBER: US/09/964,824
;; PRIOR FILING DATE: 2001-09-27
;; PRIOR APPLICATION NUMBER: US/09/967,768
;; PRIOR FILING DATE: 2001-09-28
;; PRIOR APPLICATION NUMBER: US/09/968,007
;; PRIOR FILING DATE: 2001-10-02
;; PRIOR APPLICATION NUMBER: US/09/969,347
;; PRIOR FILING DATE: 2001-10-02
;; PRIOR APPLICATION NUMBER: US/09/969,708
;; PRIOR FILING DATE: 2001-10-03
;; Remaining Prior Application data removed - See File Wrapper or PALM.
```

```
;; NUMBER OF SEQ ID NOS: 8447
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 4756
;; LENGTH: 348
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(348)
;; OTHER INFORMATION: n=a,t,g or c
US-10-843-641A-4756
```

```
Query Match          79.2%; Score 19.8; DB 9; Length 348;
```

```
Best Local Similarity 91.3%; Pred. No. 5.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGTCCCAATTGCA 23
Db      103 AAAAAATTAAGCCCAATTGCA 81
```

RESULT 29

```
US-10-027-632-324607
;; Sequence 324607, Application US/10027632
;; Publication No. US20020198371A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 324607
;; LENGTH: 463
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-324607
```

```
Query Match          79.2%; Score 19.8; DB 5; Length 483;
Best Local Similarity 91.3%; Pred. No. 5.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGTCCCAATTGCA 23
Db      251 AAAAAATTAAGTCCCAATTGCA 273
```

RESULT 30

```
US-10-027-632-324607
;; Sequence 324607, Application US/10027632
;; Publication No. US20030204075A9
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
```

```
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324607
; LENGTH: 483
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324607

Query Match          79.2%; Score 19.8; DB 6; Length 483;
Best Local Similarity 91.3%; Pred. No. 5.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 23
Db 251 AAAAAAAAAATCCCAATTCACA 273

RESULT 31
US-10-027-632-82159
; Sequence 82159, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82159
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-82159

Query Match          79.2%; Score 19.8; DB 5; Length 495;
Best Local Similarity 91.3%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 174 AAAAAAAAAATCCCAATTCAGAT 196

RESULT 32
US-10-027-632-302138
; Sequence 302138, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
```

```
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 302138
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-302138

Query Match          79.2%; Score 19.8; DB 6; Length 495;
Best Local Similarity 91.3%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 174 AAAAAAAAAATCCCAATTCAGAT 196

RESULT 33
US-10-027-632-82159
; Sequence 82159, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82159
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-82159

Query Match          79.2%; Score 19.8; DB 6; Length 495;
Best Local Similarity 91.3%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 174 AAAAAAAAAATCCCAATTCAGAT 196

RESULT 34
```



```
US-10-027-632-302138
; Sequence 302138, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 302138
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-302138

Query Match          79.2%; Score 19.8; DB 6; Length 495;
Best Local Similarity 91.3%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAAAAAAAAAGTCCCAATTCAGAT 24
      ||||||| ||||| |||||
Db      174 AAAAAAAAAATCCCAATTCAGAT 196

RESULT 35
US-09-925-065A-338895/C
; Sequence 338895, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 338895
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-338895

Query Match          79.2%; Score 19.8; DB 4; Length 512;
Best Local Similarity 91.3%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGTCCCAATTCAGA 23
      ||||||| ||||| |||||
Db      310 AAAAAAAAAAGGCCCAATTCAGA 288

RESULT 36
US-09-925-065A-338896/C
; Sequence 338896, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 338896
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-338896

Query Match          79.2%; Score 19.8; DB 4; Length 512;
Best Local Similarity 91.3%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTCCCAATTCAGA 23
      ||||||| ||||| |||||
Db      310 AAAAAAAAAAGGCCCAATTCAGA 288

RESULT 37
US-09-925-065A-389484
; Sequence 389484, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 389484
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-389484
```

Query Match 79.2%; Score 19.8; DB 4; Length 586;
Best Local Similarity 91.3%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGT 24
Db 387 AAAAAAAAAAGTCCCAATTCAGT 409

RESULT 38

US-09-925-065A-375699
; Sequence 375699, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 375699
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-375699

Query Match 79.2%; Score 19.8; DB 4; Length 616;
Best Local Similarity 91.3%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 283 AAAAAAAAAAGTCCCAATTCACA 305

RESULT 39

US-09-925-065A-510209
; Sequence 510209, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 510209
; LENGTH: 668
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-925-065A-510209

Query Match 79.2%; Score 19.8; DB 4; Length 668;
Best Local Similarity 84.0%; Pred. No. 5.6e+02;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 184 AAAAAAAAAATCTTATTCARATA 208

RESULT 40

US-09-925-065A-51386/c
; Sequence 51386, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51386
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-51386

Query Match 79.2%; Score 19.8; DB 4; Length 1662;
Best Local Similarity 91.3%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 1461 AAAAAAAAAAGTCCCAATTCAGA 1439

RESULT 41

US-09-925-065A-51387/c
; Sequence 51387, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 51387
LENGTH: 1662
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-51387

Query Match 79.2%; Score 19.8; DB 4; Length 1662;
Best Local Similarity 91.3%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGCA 23
DB 1461 AAAAAATTAAGCCCAATTGCA 1439

RESULT 42
US-10-723-860-2033
Sequence 2033, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
APPLICANT: Zlotnik, Albert
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
FILE REFERENCE: 05882.0193.NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2033
LENGTH: 6974
TYPE: DNA
ORGANISM: Homo sapiens
US-10-723-860-2033

Query Match 79.2%; Score 19.8; DB 8; Length 6974;
Best Local Similarity 91.3%; Pred. No. 7.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGCA 23
DB 2822 AAAAAACAAAGTCCCAATTAGA 2844

RESULT 43
US-09-974-298-124
Sequence 124, Application US/09974298
Patent No. US20020156263A1
GENERAL INFORMATION:
APPLICANT: Chen, Huel-Mei
TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
FILE REFERENCE: PA-0037 P
CURRENT APPLICATION NUMBER: US/09/974,298
CURRENT FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: 60/238,331
PRIOR FILING DATE: 2000-05-10
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 124
LENGTH: 7032
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020156263A1 474916.2
LOCATION: 3880-3917
OTHER INFORMATION: a, t, c, g, or other
US-09-974-298-124

Query Match 79.2%; Score 19.8; DB 3; Length 7032;
Best Local Similarity 91.3%; Pred. No. 7.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGCA 23
DB 2875 AAAAAACAAAGTCCCAATTAGA 2897

RESULT 44
US-10-723-860-6323
Sequence 6323, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
APPLICANT: Zlotnik, Albert
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
FILE REFERENCE: 05882.0193.NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6323
LENGTH: 7835
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (7174)..(7194)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (7513)..(7575)
OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-6323

Query Match 79.2%; Score 19.8; DB 8; Length 7835;
Best Local Similarity 91.3%; Pred. No. 7.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGCA 23
DB 2896 AAAAAACAAAGTCCCAATTAGA 2918

RESULT 45
US-09-764-877-3470
Sequence 3470, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et. al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3470
LENGTH: 10514
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-3470

Query Match 79.2%; Score 19.8; DB 3; Length 10514;
Best Local Similarity 91.3%; Pred. No. 8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGCA 23

Db 6354 AAAAAAAGTCCCAATTAGA 6376

RESULT 46
US-10-242-515-3470

Sequence 3470, Application US/10242515
Publication No. US2004009488A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005C1
CURRENT APPLICATION NUMBER: US/10/242,515
CURRENT FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/764,877
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/218,290
PRIOR FILING DATE: 2000-07-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3470
LENGTH: 10514
TYPE: DNA
ORGANISM: Homo sapiens
US-10-242-515-3470

Query Match 79.2%; Score 19.8; DB 6; Length 10514;
Best Local Similarity 91.3%; Pred. No. 8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAGTCCCAATTAGA 23
Db 6354 AAAAAAAGTCCCAATTAGA 6376

RESULT 47
US-10-681-773-134

Sequence 134, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
CURRENT FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 134
LENGTH: 25

TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-134

Query Match 77.6%; Score 19.4; DB 7; Length 25;
Best Local Similarity 95.2%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 AAAAAAGTCCCAATTCAGATA 25
Db 1 AAAAAAGTCCCAATTCAGATA 21

RESULT 48
US-09-925-065A-574492/c

Sequence 574492, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 574492
LENGTH: 532
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-574492

Query Match 77.6%; Score 19.4; DB 4; Length 532;
Best Local Similarity 95.2%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAGTCCCAATTCA 21
Db 334 AAAAAAAGTCCCAATTCA 314

RESULT 49
US-09-925-065A-574493/c

Sequence 574493, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 574493
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-574493

```

```

Query Match
Best Local Similarity 77.6%; Score 19.4; DB 4; Length 532;
Matches 20; Conservative 0; Pred. No. 7.8e+02; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAAAGTCCCAATTCA 21
Db 334 AAAAAAAAAAGTCCCAATTCA 314

```

```

RESULT 50
US-09-925-065A-394568/C
; Sequence 394568, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 394568
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-394568

```

```

Query Match
Best Local Similarity 77.6%; Score 19.4; DB 4; Length 569;
Matches 20; Conservative 0; Pred. No. 7.8e+02; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAAAGTCCCAATTCA 21
Db 241 AAAAAAAAAAGTCCCAATTCA 221

```

Search completed: December 14, 2005, 08:46:37
 Job time : 377.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 : Search time 373.2 Seconds
(without alignments)
553.951 Million cell updates/sec

Title: US-10-681-773-8
Perfect score: 25
Sequence: 1 aaaaaaaaaagttccaatcagata 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Published Applications NA_Main.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	US-10-681-773-8	Sequence 8, Appli
2	24.6	98.4	33	US-10-891-260-5923	Sequence 5923, Ap
3	24.6	98.4	425	US-09-925-065A-169174	Sequence 169174,
4	24	96.0	25	US-10-681-773-15159	Sequence 15, Appl
5	24	96.0	25	US-10-681-773-62159	Sequence 62159, A
6	23.4	93.6	25	US-10-681-773-7	Sequence 7, Appli
7	23.4	93.6	418	US-09-925-065A-169175	Sequence 169175,
8	23	92.0	25	US-10-681-773-14	Sequence 14, Appl
9	22.4	89.6	25	US-10-681-773-14	Sequence 14, Appl
10	22.4	89.6	25	US-10-681-773-62158	Sequence 62158, A
11	21.8	87.6	1270	US-09-925-065A-691066	Sequence 691066,
12	21.8	87.6	25	US-10-681-773-18418	Sequence 18418, A
13	21	83.0	25	US-10-681-773-134	Sequence 134, App
14	20.8	83.2	326	US-10-424-599-134659	Sequence 134659,
15	20.8	83.2	546	US-09-925-065A-544535	Sequence 544535,
16	20.8	83.2	546	US-09-925-065A-544535	Sequence 544535,
17	20.8	83.2	584	US-10-425-115-33678	Sequence 33678, A
18	20.8	83.2	2567	US-10-425-114-29775	Sequence 29775, A
19	20.8	83.2	3310	US-10-424-599-109133	Sequence 109133,
20	20.8	83.2	31271	US-11-097-143-37261	Sequence 37261, A
21	20.8	83.2	121410	US-10-741-600-41773	Sequence 41773, A
22	20.4	81.6	580	US-09-925-065A-660788	Sequence 660788,
23	20.2	80.8	265	US-10-425-115-55912	Sequence 55912, A

24	20.2	80.8	450	US-10-027-632-292613	Sequence 292613,
25	20.2	80.8	450	US-10-027-632-292613	Sequence 292613,
26	20.2	80.8	2839	US-10-424-599-138849	Sequence 138849,
27	20.2	80.8	3759	US-10-270-333-28	Sequence 28, Appl
28	20.2	80.8	3759	US-11-097-143-10693	Sequence 10693, A
29	20.2	80.8	4159	US-10-097-143-22276	Sequence 22276, A
30	20.2	80.8	5933	US-10-257-166-59	Sequence 59, Appl
31	20.2	80.8	5933	US-10-257-166-60	Sequence 60, Appl
32	20.2	80.8	6486	US-10-433-793-80	Sequence 80, Appl
33	20.2	80.8	9510	US-10-221-174-1160	Sequence 116, App
34	20.2	80.8	9510	US-10-221-174-1160	Sequence 160, App
35	20.2	80.8	10151	US-10-311-455-2405	Sequence 2405, App
36	20.2	80.8	10151	US-10-311-455-2405	Sequence 2406, App
37	20.2	80.8	12542	US-09-764-664-1774	Sequence 1774, App
38	20.2	80.8	18750	US-10-719-993-6900	Sequence 6900, App
39	20.2	80.8	20142	US-10-282-122A-7627	Sequence 7627, App
40	20.2	80.8	109586	US-10-087-192-220	Sequence 220, App
41	20.2	80.8	398287	US-10-741-601-5719	Sequence 5719, App
42	20.2	80.8	398287	US-10-741-601-5719	Sequence 17839, A
43	19.4	77.6	25	US-10-681-773-131	Sequence 131, App
44	19.4	77.6	667	US-10-027-632-253929	Sequence 253929,
45	19.4	77.6	1146	US-10-027-632-251319	Sequence 251319,
46	19.4	77.6	1146	US-10-027-632-251320	Sequence 251320,
47	19.4	77.6	1146	US-10-027-632-251321	Sequence 251321,
48	19.4	77.6	1146	US-10-027-632-251319	Sequence 251319,
49	19.4	77.6	1146	US-10-027-632-251320	Sequence 251320,
50	19.4	77.6	1146	US-10-027-632-251321	Sequence 251321,
51	19.4	77.6	1174	US-09-925-065A-1118	Sequence 1118, Ap
52	19.2	76.8	201	US-10-741-600-34144	Sequence 34144, A
53	19.2	76.8	201	US-10-741-600-66816	Sequence 66816, A
54	19.2	76.8	424	US-09-925-065A-345806	Sequence 345806,
55	19.2	76.8	424	US-09-925-065A-345807	Sequence 345807,
56	19.2	76.8	424	US-09-925-065A-345808	Sequence 345808,
57	19.2	76.8	424	US-09-925-065A-345809	Sequence 345809,
58	19.2	76.8	424	US-09-925-065A-345810	Sequence 345810,
59	19.2	76.8	521	US-09-925-065A-126942	Sequence 126942,
60	19.2	76.8	530	US-09-925-065A-495622	Sequence 495622,
61	19.2	76.8	530	US-09-925-065A-495623	Sequence 495623,
62	19.2	76.8	532	US-09-925-065A-495621	Sequence 495621,
63	19.2	76.8	533	US-10-357-930-10712	Sequence 10712, A
64	19.2	76.8	535	US-09-925-065A-425519	Sequence 425519,
65	19.2	76.8	535	US-09-925-065A-425520	Sequence 425520,
66	19.2	76.8	535	US-09-925-065A-425521	Sequence 425521,
67	19.2	76.8	538	US-09-925-065A-927714	Sequence 927714,
68	19.2	76.8	561	US-09-925-065A-743895	Sequence 743895,
69	19.2	76.8	571	US-10-425-115-151846	Sequence 151846,
70	19.2	76.8	583	US-09-925-065A-411648	Sequence 411648,
71	19.2	76.8	599	US-10-972-079-89234	Sequence 89234, A
72	19.2	76.8	600	US-09-925-065A-92089	Sequence 92089, A
73	19.2	76.8	600	US-09-925-065A-92090	Sequence 92090, A
74	19.2	76.8	605	US-09-925-065A-381588	Sequence 381588,
75	19.2	76.8	754	US-10-027-632-133741	Sequence 133741,
76	19.2	76.8	754	US-10-027-632-133741	Sequence 133741,
77	19.2	76.8	822	US-10-357-930-11351	Sequence 11351, A
78	19.2	76.8	968	US-09-925-065A-711808	Sequence 717808,
79	19.2	76.8	3012	US-11-097-143-39937	Sequence 39937, A
80	19.2	76.8	3450	US-10-052-589-3	Sequence 3, Appli
81	19.2	76.8	6392	US-10-311-455-657	Sequence 59, Appl
82	19.2	76.8	6392	US-10-311-455-657	Sequence 60, Appl
83	19.2	76.8	9707	US-10-311-455-1394	Sequence 1394, App
84	19.2	76.8	32221	US-09-764-647-1406	Sequence 1406, App
85	19.2	76.8	32221	US-10-092-154-1406	Sequence 1406, App
86	19.2	76.8	111330	US-10-741-600-17920	Sequence 17920, A
87	19.2	76.8	111330	US-10-741-600-17920	Sequence 590, App
88	19.2	76.8	274869	US-10-741-600-176580	Sequence 176580, A
89	18.8	75.2	4400	US-10-027-632-182884	Sequence 182884,
90	18.8	75.2	450	US-10-027-632-182884	Sequence 182884,
91	18.8	75.2	521	US-09-925-065A-182884	Sequence 182884,
92	18.8	75.2	600	US-10-972-079-47029	Sequence 47029, A
93	18.8	75.2	600	US-10-972-079-47030	Sequence 47030, A
94	18.8	75.2	841	US-10-027-632-172584	Sequence 172584,
95	18.8	75.2	841	US-10-027-632-172584	Sequence 172584,
96	18.8	75.2	915	US-10-027-632-155121	Sequence 155121,

97 18.8 75.2 915 6 US-10-027-632-155121
C 98 18.8 75.2 2762 4 US-09-925-065A-676734
99 18.6 74.4 102 2 US-08-781-986A-1391
100 18.6 74.4 102 7 US-10-329-624-1391
101 18.6 74.4 201 7 US-10-741-601-23277
C 102 18.6 74.4 201 8 US-10-719-993-30903
C 103 18.6 74.4 201 8 US-10-719-993-31238
C 104 18.6 74.4 201 8 US-10-719-993-31294
C 105 18.6 74.4 201 8 US-10-719-993-31296
C 106 18.6 74.4 226 8 US-10-425-115-103214
C 107 18.6 74.4 265 2 US-08-781-986A-1119
C 108 18.6 74.4 265 7 US-10-329-624-1119
C 109 18.6 74.4 291 7 US-10-424-599-102609
C 110 18.6 74.4 300 9 US-10-779-543-2779
C 111 18.6 74.4 319 7 US-10-437-963-18768
C 112 18.6 74.4 322 8 US-10-674-124A-15033
C 113 18.6 74.4 328 7 US-10-424-599-105407
C 114 18.6 74.4 329 7 US-10-424-599-17540
C 115 18.6 74.4 353 4 US-09-925-065A-155118
C 116 18.6 74.4 353 4 US-09-925-065A-578830
C 117 18.6 74.4 370 7 US-10-424-599-90033
C 118 18.6 74.4 370 8 US-10-674-124A-11124
C 119 18.6 74.4 371 7 US-10-424-599-136227
C 120 18.6 74.4 384 8 US-10-425-115-75472
C 121 18.6 74.4 412 7 US-10-424-599-90005
C 122 18.6 74.4 417 8 US-10-357-930-23469
C 123 18.6 74.4 417 8 US-10-357-930-29356
C 124 18.6 74.4 419 7 US-10-424-599-105846
C 125 18.6 74.4 422 8 US-10-357-930-14663
C 126 18.6 74.4 432 7 US-10-424-599-81929
C 127 18.6 74.4 433 5 US-10-027-632-62913
C 128 18.6 74.4 439 5 US-10-027-632-62914
C 129 18.6 74.4 439 5 US-10-027-632-64268
C 130 18.6 74.4 439 5 US-10-027-632-297404
C 131 18.6 74.4 439 5 US-10-027-632-297405
C 132 18.6 74.4 439 6 US-10-027-632-62913
C 133 18.6 74.4 439 6 US-10-027-632-62914
C 134 18.6 74.4 439 6 US-10-027-632-64268
C 135 18.6 74.4 439 6 US-10-027-632-297404
C 136 18.6 74.4 439 6 US-10-027-632-297405
C 137 18.6 74.4 451 8 US-10-357-930-5494
C 138 18.6 74.4 463 8 US-10-357-930-35768
C 139 18.6 74.4 463 8 US-10-357-930-4572
C 140 18.6 74.4 470 4 US-09-925-065A-701169
C 141 18.6 74.4 474 4 US-09-925-065A-285158
C 142 18.6 74.4 479 4 US-09-925-065A-734121
C 143 18.6 74.4 479 4 US-09-925-065A-812661
C 144 18.6 74.4 500 3 US-09-991-936-86
C 145 18.6 74.4 500 9 US-10-978-245-86
C 146 18.6 74.4 504 4 US-09-925-065A-477576
C 147 18.6 74.4 510 4 US-09-925-065A-629394
C 148 18.6 74.4 510 5 US-10-027-632-35279
C 149 18.6 74.4 510 6 US-10-027-632-35279
C 150 18.6 74.4 514 4 US-09-925-065A-268873

ALIGNMENTS

RESULT 1
US-10-681-773-8
; Sequence 8, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475

Sequence 155121,
Sequence 676734,
Sequence 1391, Ap
Sequence 1391, Ap
Sequence 23277, A
Sequence 30903, A
Sequence 31238, A
Sequence 31294, A
Sequence 31296, A
Sequence 103214,
Sequence 1119, Ap
Sequence 1119, Ap
Sequence 102609,
Sequence 2779, Ap
Sequence 18768, A
Sequence 15033, A
Sequence 105407,
Sequence 17540, A
Sequence 155118,
Sequence 578830,
Sequence 90033, A
Sequence 11124, A
Sequence 136227,
Sequence 75472, A
Sequence 90005, A
Sequence 23469, A
Sequence 23566, A
Sequence 105846,
Sequence 14663, A
Sequence 81929, A
Sequence 62913, A
Sequence 62914, A
Sequence 64268, A
Sequence 297404, A
Sequence 297405,
Sequence 62913, A
Sequence 62914, A
Sequence 64268, A
Sequence 297404, A
Sequence 297405,
Sequence 5494, Ap
Sequence 35768, A
Sequence 4572, A
Sequence 701169,
Sequence 285158,
Sequence 734121,
Sequence 812661,
Sequence 86, Appl
Sequence 86, Appl
Sequence 477576,
Sequence 629394,
Sequence 35279, A
Sequence 35279, A
Sequence 268873,

; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 8
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-8

Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTGAGATA 25
Db 1 AAAAAAAAAAGTCCATTGAGATA 25

RESULT 2
US-10-691-260-5923
; Sequence 5923, Application US/10891260
; Publication No. US20050227244A1
; GENERAL INFORMATION:
; APPLICANT: Affymetrix, Inc.
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; TITLE OF INVENTION: Methods of Analysis of Human Polymorphisms
; FILE REFERENCE: 3522.3
; CURRENT APPLICATION NUMBER: US/10/891,260
; CURRENT FILING DATE: 2004-07-13
; PRIOR APPLICATION NUMBER: 10/681,773
; PRIOR FILING DATE: 2003-10-07
; NUMBER OF SEQ ID NOS: 10244
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5923
; LENGTH: 33
; TYPE: DNA
; ORGANISM: homo sapien
US-10-691-260-5923

Query Match 98.4%; Score 24.6; DB 9; Length 33;
Best Local Similarity 96.0%; Pred. No. 9;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTGAGATA 25
Db 5 AAAAAAAAAAGTCCATTGAGATA 29

RESULT 3
US-09-925-065A-169174
; Sequence 169174, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957066

US-09-925-065A-169174
; Sequence 169174, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957066


```
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169174
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-169174

Query Match
Best Local Similarity 96.0%; Score 24.6; DB 4; Length 425;
Best Local Similarity 96.0%; Pred. No. 13;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTCCAGATA 25
Db 120 AAAAAAAAAAGTCCATTCCAGATA 144

RESULT 4
US-10-681-773-15
; Sequence 15, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 15
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-15

Query Match
Best Local Similarity 96.0%; Score 24; DB 7; Length 25;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCATTCCAGATA 25
Db 1 AAAAAAAAAAGTCCATTCCAGATA 24

RESULT 5
US-10-681-773-62159
; Sequence 62159, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 62159
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
```

```
US-10-681-773-62159

Query Match
Best Local Similarity 96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTCCAGAT 24
Db 2 AAAAAAAAAAGTCCATTCCAGAT 25

RESULT 6
US-10-681-773-7
; Sequence 7, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 7
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-7

Query Match
Best Local Similarity 93.6%; Score 23.4; DB 7; Length 25;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTCCAGATA 25
Db 1 AAAAAAAAAAGTCCATTCCAGATA 25

RESULT 7
US-09-925-065A-169175
; Sequence 169175, Application US/09925065A
; Publication No. US2005028172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169175
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-169175
```

Query Match 93.6%; Score 23.4; DB 4; Length 418;
Best Local Similarity 96.0%; Pred. No. 36;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 115 AAAAAAAAAAGTTCATTCAGATA 139

RESULT 8

US-10-681-773-18419
Sequence 18419, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
PRIOR FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 18419
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-18419

Query Match 92.0%; Score 23; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGA 23
Db 3 AAAAAAAAAAGTTCATTCAGA 25

RESULT 9

US-10-681-773-14
Sequence 14, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
PRIOR FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 14
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-14

Query Match 89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 60;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTTCATTCAGATA 25
Db 1 AAAAAAAAAAGTTCATTCAGATA 24

RESULT 10
US-10-681-773-62158
Sequence 62158, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
PRIOR FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 62158
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-62158

Query Match 89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 60;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 2 AAAAAAAAAAGTTCATTCAGAT 25

RESULT 11
US-09-925-065A-691066/C
Sequence 691066, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 691066
LENGTH: 1270
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711,
LOCATION: 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723,
LOCATION: 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735,
LOCATION: 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746
OTHER INFORMATION: n = A,T,C or G

```
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758,
; LOCATION: 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770,
; LOCATION: 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782,
; LOCATION: 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793
; OTHER INFORMATION: n = A,T,C or G
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 794, 795, 796, 797, 798, 799
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-691066

Query Match      87.2%; Score 21.8; DB 4; Length 1270;
Best Local Similarity 92.0%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 1146 AAAAAAAAAATTCCATTTCAGATA 1122

RESULT 12
US-10-681-773-18418
; Sequence 18418, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 18418
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-18418

Query Match      85.6%; Score 21.4; DB 7; Length 25;
Best Local Similarity 95.7%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTTCAGA 23
Db 3 AAAAAAAAAAGTCCATTTCAGA 25

RESULT 13
US-10-681-773-134
; Sequence 134, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08

NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 134
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-134

Query Match      84.0%; Score 21; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AAAAAAAAAAGTCCATTTCAGATA 25
Db 1 AAAAAAAAAAGTCCATTTCAGATA 21

RESULT 14
US-10-424-599-134659/c
; Sequence 134659, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 134659
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92606C.1
US-10-424-599-134659

Query Match      83.2%; Score 20.8; DB 7; Length 326;
Best Local Similarity 91.7%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCATTTCAGATA 25
Db 316 AAAAAAAAAATTCCATTTCATATA 293

RESULT 15
US-09-925-065A-793833/c
; Sequence 793833, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 793833
```

```
; LENGTH: 524
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-793833

Query Match      83.2%; Score 20.8; DB 4; Length 524;
Best Local Similarity 91.7%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 224 AAAAAAAAAAGTCCCAATTCAGAT 201

RESULT 16
US-09-925-065A-544535/c
; Sequence 544535, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIORITY FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 544535
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-544535

Query Match      83.2%; Score 20.8; DB 4; Length 546;
Best Local Similarity 91.7%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 402 AAAAAAAAAAGTCCCAATTCAGAT 379

RESULT 17
US-10-425-115-33678/c
; Sequence 33678, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425.115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 33678
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
```

```
; LOCATION: (1)..(584)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_13071C.1
US-10-425-115-33678

Query Match      83.2%; Score 20.8; DB 8; Length 584;
Best Local Similarity 91.7%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 291 AAAAAAAAAAGTCCCAATTCAGAT 268

RESULT 18
US-10-425-114-29775/c
; Sequence 29775, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven B.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425.114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29775
; LENGTH: 2567
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMNONIR031D10_FLI
US-10-425-114-29775

Query Match      83.2%; Score 20.8; DB 7; Length 2567;
Best Local Similarity 91.7%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 2388 AAAAAAAAAAGTCCCAATTCAGATA 2365

RESULT 19
US-10-424-599-109133/c
; Sequence 109133, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 109133
; LENGTH: 3310
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_69560C.1
US-10-424-599-109133

Query Match      83.2%; Score 20.8; DB 7; Length 3310;
```

Best Local Similarity 91.7%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTTCATTCAGATA 25
Db 3018 AAAAAAAAAAGTTCATTCAGATA 2995

RESULT 20

US-11-097-143-37261
; Sequence 37261, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: C1000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37261
; LENGTH: 31271
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-37261

Query Match 83.2%; Score 20.8; DB 10; Length 31271;
Best Local Similarity 91.7%; Pred. No. 6.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 6323 AAAAAAAAAAGTTCATTCAGAT 6346

RESULT 21
US-10-741-600-17733/c
; Sequence 17733, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C100149
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17733
; LENGTH: 121410
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-17733

Query Match 83.2%; Score 20.8; DB 8; Length 121410;
Best Local Similarity 91.7%; Pred. No. 7.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 32408 AAAAAAAAAAGTTCATTCAGAT 32385

RESULT 22

US-09-925-065A-660788
; Sequence 660788, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 660788
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-660788

Query Match 81.6%; Score 20.4; DB 4; Length 580;
Best Local Similarity 95.5%; Pred. No. 5.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAG 22
Db 40 AAAAAAAAAAGTTCATTCAG 61

RESULT 23
US-10-425-115-55912/c
; Sequence 55912, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovallik, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 55912
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_150986C.1
US-10-425-115-55912

Query Match 80.8%; Score 20.2; DB 8; Length 265;
Best Local Similarity 88.0%; Pred. No. 5.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy 1 AAAAAAAAAAGTTCATTGAGATA 25
| | | | | | | | | | | | | | | |
Db 60 AAAAAAAAAATTCCATTCAAAA 36

RESULT 24
US-10-027-632-292613
; Sequence 292613, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 292613
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-292613

Query Match 80.8%; Score 20.2; DB 5; Length 450;
Best Local Similarity 88.0%; Pred. No. 6.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTGAGATA 25
| | | | | | | | | | | | | | | |
Db 59 AAAAAAAAAAGTTCCAATAAATA 83

RESULT 25
US-10-027-632-292613
; Sequence 292613, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
```

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 292613
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-292613

Query Match 80.8%; Score 20.2; DB 6; Length 450;
Best Local Similarity 88.0%; Pred. No. 6.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTGAGATA 25
| | | | | | | | | | | | | | | |
Db 59 AAAAAAAAAAGTTCCAATAAATA 83

RESULT 26
US-10-424-599-138849/c
; Sequence 138849, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovall, David K
; APPLICANT: Zhou Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 138849
; LENGTH: 2839
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_96388C.1
US-10-424-599-138849

Query Match 80.8%; Score 20.2; DB 7; Length 2839;
Best Local Similarity 88.0%; Pred. No. 7.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTGAGATA 25
| | | | | | | | | | | | | | | |
Db 2787 AAAAAAAAAAGTTCATTCTAGTA 2763

RESULT 27
US-10-270-333-28/c
; Sequence 28, Application US/10270333
; Publication No. US20030092124A1
; GENERAL INFORMATION:
; APPLICANT: Ciavichk, Anibal
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS, AND USES
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR, PROTEINS, AND USES
; FILE REFERENCE: CI000733CON
; CURRENT APPLICATION NUMBER: US/10/270,333
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/168,677
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/175,691
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/191,638
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 3759
; TYPE: DNA
; ORGANISM: Drosophila
```

US-10-270-333-28

Query Match 80.8%; Score 20.2; DB 5; Length 3759;
Best Local Similarity 88.0%; Pred. No. 8.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATTCATTGAGATA 25
Db 303 AAAAAAAAACTTACATTGAGATA 279

RESULT 28

US-11-097-143-10693/c

; Sequence 10693, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10693
; LENGTH: 3759
; TYPE: DNA
; ORGANISM: DROSOPHILA
; US-11-097-143-10693

Query Match 80.8%; Score 20.2; DB 10; Length 3759;
Best Local Similarity 88.0%; Pred. No. 8.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATTCATTGAGATA 25
Db 303 AAAAAAAAACTTACATTGAGATA 279

RESULT 29

US-11-097-143-22276/c

; Sequence 22276, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191

; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22276
; LENGTH: 4169
; TYPE: DNA
; ORGANISM: DROSOPHILA
; US-11-097-143-22276

Query Match 80.8%; Score 20.2; DB 10; Length 4169;
Best Local Similarity 88.0%; Pred. No. 8.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATTCATTGAGATA 25
Db 3426 AAAAAAAAACTTACATTGAGATA 3402

RESULT 30

US-10-257-166-59/c

; Sequence 59, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERROCK, Christian
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; PRIOR FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 59
; LENGTH: 5933
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-257-166-59

Query Match 80.8%; Score 20.2; DB 7; Length 5933;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATTCATTGAGATA 25
Db 2331 AAAAAAAAAATTCATTGAGATA 2307

RESULT 31

US-10-257-166-60

; Sequence 60, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander

```
/ APPLICANT: PIEPENBROCK, Christian
/ APPLICANT: BERLIN, Kurt
/ TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
/ TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
/ FILE REFERENCE: 5013.1011
/ CURRENT APPLICATION NUMBER: US/10/257,166
/ PRIOR FILING DATE: 2002-10-07
/ PRIOR APPLICATION NUMBER: PCT/EP01/07470
/ DE 10032529.7
/ DE 10043826.1
/ PRIOR FILING DATE: 2001-06-29
/ 2000-06-30
/ 2000-09-01
/ NUMBER OF SEQ ID NOS: 178
/ SEQ ID NO 60
/ LENGTH: 5933
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-60

Query Match      80.8%; Score 20.2; DB 7; Length 5933;
Best Local Similarity 88.0%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTGAGATA 25
Db 3603 AAAAAAAAAAGTTTAATTGATA 3627

RESULT 32
US-10-433-793-80/c
/ Sequence 80, Application US/10433793
/ Publication No. US200401423341
/ GENERAL INFORMATION:
/ APPLICANT: Epigenomics AG
/ TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
/ FILE REFERENCE:
/ CURRENT APPLICATION NUMBER: US/10/433,793
/ CURRENT FILING DATE: 2003-06-06
/ NUMBER OF SEQ ID NOS: 212
/ SEQ ID NO 80
/ LENGTH: 6486
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-80

Query Match      80.8%; Score 20.2; DB 7; Length 6486;
Best Local Similarity 88.0%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTGAGATA 25
Db 6080 AAAAAAAAAATTCAATTGACATA 6056

RESULT 33
US-10-240-485-116/c
/ Sequence 116, Application US/10240485
/ Publication No. US20030148327A1
/ GENERAL INFORMATION:
/ APPLICANT: OLEK, Alexander
/ APPLICANT: PIEPENBROCK, Christian
/ APPLICANT: BERLIN, Kurt
/ TITLE OF INVENTION: Diagnosis of Diseases Associated with
/ TITLE OF INVENTION: Metastasis
/ FILE REFERENCE: 5013.1007
/ CURRENT APPLICATION NUMBER: US/10/240,485
/ CURRENT FILING DATE: 2002-10-02
/ PRIOR APPLICATION NUMBER: PCT/EP01/03970
```

```
/ PRIOR FILING DATE: 2001-04-06
/ PRIOR APPLICATION NUMBER: DE 10019058.8
/ PRIOR FILING DATE: 2000-04-06
/ PRIOR APPLICATION NUMBER: DE 10019173.8
/ PRIOR FILING DATE: 2000-04-07
/ PRIOR APPLICATION NUMBER: DE 10032529.7
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: DE 10043826.1
/ PRIOR FILING DATE: 2000-09-01
/ NUMBER OF SEQ ID NOS: 202
/ SEQ ID NO 116
/ LENGTH: 9510
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (695, 710, 822, 1025, 1027, 1639, 1677, 1732, 1999, 2010, 2054)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (2066, 2660, 2675, 2813, 2817, 3198, 3407, 4097, 4239, 4250)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (4673, 4706, 4775, 4784, 5114, 5308, 5350, 5655, 5669, 5804)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (5821, 6088, 6568, 6580, 6588, 7576, 7988, 8038, 8129, 8133)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (8323, 8328, 8394, 8732, 8765, 8819, 8885, 9201)
US-10-240-485-116

Query Match      80.8%; Score 20.2; DB 6; Length 9510;
Best Local Similarity 88.0%; Pred. No. 9.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTGAGATA 25
Db 5970 AAAAAAAAAATTCAATTAAATA 5946

RESULT 34
US-10-221-714A-160/c
/ Sequence 160, Application US/10221714A
/ Publication No. US2004048254A1
/ GENERAL INFORMATION:
/ APPLICANT: OLEK, Alexander
/ APPLICANT: PIEPENBROCK, Christian
/ APPLICANT: BERLIN, Kurt
/ TITLE OF INVENTION: Diagnosis of Diseases Associated with
/ TITLE OF INVENTION: Tumor suppressor genes and oncogenes
/ FILE REFERENCE: 5013.1005
/ CURRENT APPLICATION NUMBER: US/10/221,714A
/ CURRENT FILING DATE: 2003-01-21
/ PRIOR APPLICATION NUMBER: PCT/EP01/02955
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: DE 10013847.0
/ PRIOR FILING DATE: 2000-03-15
/ PRIOR APPLICATION NUMBER: DE 10019058.8
/ PRIOR FILING DATE: 2000-04-06
/ PRIOR APPLICATION NUMBER: DE 10019173.8
/ PRIOR FILING DATE: 2000-04-07
/ PRIOR APPLICATION NUMBER: DE 10032529.7
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: DE 10043826.1
/ PRIOR FILING DATE: 2000-09-01
/ NUMBER OF SEQ ID NOS: 540
/ SEQ ID NO 160
/ LENGTH: 9510
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
```



```
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: 695, 710, 822, 1025, 1027, 1639, 1677, 1732, 1999, 2010, 2054
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 2066, 2660, 2675, 2813, 2817, 3198, 3407, 4097, 4239, 4250
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 4673, 4706, 4775, 4784, 5114, 5308, 5330, 5655, 5669, 5804
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 5821, 6088, 6568, 6580, 6588, 7576, 7988, 8038, 8129, 8133
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 8323, 8328, 8394, 8732, 8765, 8819, 8885, 9201
OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-160

Query Match      80.8%; Score 20.2; DB 7; Length 9510;
Best Local Similarity 88.0%; Pred. No. 9.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 5970 AAAAAAAAAATTCATTAATAA 5946

RESULT 35
US-10-311-455-2405
; Sequence 2405, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of diseases Associated with the Immune System by Detect
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2405
; LENGTH: 10151
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: 600, 657
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-2405

Query Match      80.8%; Score 20.2; DB 6; Length 10151;
Best Local Similarity 88.0%; Pred. No. 9.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 7821 AAAAAAAAAAGTTTAATTAGATA 7845

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: 695, 710, 822, 1025, 1027, 1639, 1677, 1732, 1999, 2010, 2054
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 2066, 2660, 2675, 2813, 2817, 3198, 3407, 4097, 4239, 4250
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 4673, 4706, 4775, 4784, 5114, 5308, 5330, 5655, 5669, 5804
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 5821, 6088, 6568, 6580, 6588, 7576, 7988, 8038, 8129, 8133
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: 8323, 8328, 8394, 8732, 8765, 8819, 8885, 9201
OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-160

Query Match      80.8%; Score 20.2; DB 7; Length 9510;
Best Local Similarity 88.0%; Pred. No. 9.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 5970 AAAAAAAAAATTCATTAATAA 5946

RESULT 36
US-10-311-455-2406/c
; Sequence 2406, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2406
; LENGTH: 10151
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: 9495, 9552
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-2406

Query Match      80.8%; Score 20.2; DB 6; Length 10151;
Best Local Similarity 88.0%; Pred. No. 9.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 2331 AAAAAAAAAATTCATTAATAA 2307

RESULT 37
US-09-764-864-1774
; Sequence 1774, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN, et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764.864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1774
; LENGTH: 12542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1774

Query Match      80.8%; Score 20.2; DB 3; Length 12542;
Best Local Similarity 88.0%; Pred. No. 9.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 4239 AAAAAAAAAAGTGCAATTCAGATA 4263

RESULT 38
US-10-719-993-6900/c
; Sequence 6900, Application US/10719993
; Publication No. US20040265849A1
```

```

; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6900
; LENGTH: 18750
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-719-993-6900

Query Match      80.8%; Score 20.2; DB 8; Length 18750;
Best Local Similarity 88.0%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTTCCAATTCAGATA 25
Db      14632 AAAAAAAAAAGTTCCAATTAATA 14608

RESULT 39
US-10-282-122A-7627
; Sequence 7627, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trewick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 7627
; LENGTH: 20142
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-282-122A-7627
```

```

Query Match      80.8%; Score 20.2; DB 7; Length 20142;
Best Local Similarity 88.0%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTTCCAATTCAGATA 25
Db      2398 AAACAAACAGTTCCAATTCAGATA 2422

RESULT 40
US-10-087-192-220/c
; Sequence 220, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 220
; LENGTH: 109586
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(109586)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-220

Query Match      80.8%; Score 20.2; DB 5; Length 109586;
Best Local Similarity 88.0%; Pred. No. 1.3e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTTCCAATTCAGATA 25
Db      97181 AAAAAAAAAACAGTACATTCAGATA 97157

RESULT 41
US-10-741-601-5719/c
; Sequence 5719, Application US/10741601
; Publication No. US2004016519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5719
; LENGTH: 398287
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(398287)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-741-601-5719

Query Match      80.8%; Score 20.2; DB 7; Length 398287;
Best Local Similarity 88.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

OY 1 AAAAAAAAAAGTCCAAATTCAGATA 25
|||||
Db 229943 AAAAAAAAAAGTAAATTCATATA 229919

RESULT 42
US-10-741-600-17839/c
; Sequence 17839, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17839
; LENGTH: 398287
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(398287)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17839

Query Match 80.8%; Score 20.2; DB 8; Length 398287;
Best Local Similarity 88.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCAAATTCAGATA 25
|||||
Db 229943 AAAAAAAAAAGTAAATTCATATA 229919

RESULT 43
US-10-681-773-131
; Sequence 131, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 131
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-131

Query Match 77.6%; Score 19.4; DB 7; Length 25;
Best Local Similarity 95.2%; Pred. No. 8.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 AAAAAAGTCCAAATTCAGATA 25
|||||
Db 1 AAAAAAGTCCAAATTCAGATA 21

RESULT 44

US-10-027-632-253929
; Sequence 253929, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 253929
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-253929

Query Match 77.6%; Score 19.4; DB 5; Length 667;
Best Local Similarity 95.2%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCAAATTCATA 21
|||||
Db 483 AAAAAAAAAAGTCCAAATTTA 503

RESULT 45
US-10-027-632-253929
; Sequence 253929, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 253929
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-253929

Query Match 77.6%; Score 19.4; DB 6; Length 667;
Best Local Similarity 95.2%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTC A 21
|||||
Db 483 AAAAAAAAAAGTTCATTC A 503

RESULT 46

US-10-027-632-251319/c
; Sequence 251319, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 251319
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-251319

Query Match 77.6%; Score 19.4; DB 5; Length 1146;
Best Local Similarity 95.2%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTC A 21
|||||
Db 460 AAAAAAAAAAGTTCATTC A 440

RESULT 47

US-10-027-632-251320/c
; Sequence 251320, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 251320
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-251320

RESULT 48

Query Match 77.6%; Score 19.4; DB 5; Length 1146;
Best Local Similarity 95.2%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTC A 21
|||||
Db 460 AAAAAAAAAAGTTCATTC A 440

RESULT 49

US-10-027-632-251321/c
; Sequence 251321, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 251321
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-251321

Query Match 77.6%; Score 19.4; DB 5; Length 1146;
Best Local Similarity 95.2%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTC A 21
|||||
Db 460 AAAAAAAAAAGTTCATTC A 440

RESULT 49

US-10-027-632-251319/c
; Sequence 251319, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

Search completed: December 14, 2005, 08:46:40
Job time: 376.2 secs

```

: CURRENT FILING DATE: 2002-04-30
: PRIOR APPLICATION NUMBER: US 60/218,006
: PRIOR FILING DATE: 2000-07-12
: PRIOR APPLICATION NUMBER: US 60/198,676
: PRIOR FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: US 60/193,483
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: US 60/185,218
: PRIOR FILING DATE: 2000-02-24
: PRIOR APPLICATION NUMBER: US 60/167,363
: PRIOR FILING DATE: 1999-11-23
: PRIOR APPLICATION NUMBER: US 60/156,358
: PRIOR FILING DATE: 1999-09-28
: PRIOR APPLICATION NUMBER: US 60/146,002
: PRIOR FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 325720
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 251319
: LENGTH: 1146
: TYPE: DNA
: ORGANISM: Human
US-10-027-632-251319

```

```

Query Match      77.6%: Score 19.4; DB 6; Length 1146;
Best Local Similarity 95.2%: Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY      1 AAAAAAAAAAGTCCATGCA 21
Db      460 AAAAAAAAAAGTCCATGCA 440

```

```

RESULT 50
US-10-027-632-251320/c
: Sequence 251320, Application US/10027632
: Publication No. US20030204075A9
: GENERAL INFORMATION:
: APPLICANT: Wang, David G.
: TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
: FILE REFERENCE: 108827.129
: CURRENT APPLICATION NUMBER: US/10/027,632
: CURRENT FILING DATE: 2002-04-30
: PRIOR APPLICATION NUMBER: US 60/218,006
: PRIOR FILING DATE: 2000-07-12
: PRIOR APPLICATION NUMBER: US 60/198,676
: PRIOR FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: US 60/193,483
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: US 60/185,218
: PRIOR FILING DATE: 2000-02-24
: PRIOR APPLICATION NUMBER: US 60/167,363
: PRIOR FILING DATE: 1999-11-23
: PRIOR APPLICATION NUMBER: US 60/156,358
: PRIOR FILING DATE: 1999-09-28
: PRIOR APPLICATION NUMBER: US 60/146,002
: PRIOR FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 325720
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 251320
: LENGTH: 1146
: TYPE: DNA
: ORGANISM: Human
US-10-027-632-251320

```

```

Query Match      77.6%: Score 19.4; DB 6; Length 1146;
Best Local Similarity 95.2%: Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY      1 AAAAAAAAAAGTCCATGCA 21
Db      460 AAAAAAAAAAGTCCATGCA 440

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 ; Search time 373.2 Seconds
(without alignments)
553.951 Million cell updates/sec

Title: US-10-681-773-9

Perfect score: 25
Sequence: 1 aaaaaaaaaacatcatcaaac 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database:

Published Applications NA_Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	7	US-10-681-773-9
2	25	96.0	25	7	US-10-681-773-19
3	23.4	93.6	25	7	US-10-681-773-18
4	23	92.0	25	7	US-10-681-773-62157
5	22.4	89.6	25	7	US-10-681-773-58
6	22	88.0	25	7	US-10-681-773-18417
7	21.8	87.2	480	3	US-09-920-300A-1494
8	21.8	87.2	480	5	US-10-033-528B-1494
9	21.8	87.2	480	6	US-10-099-926-1494
10	21.8	87.2	480	9	US-10-961-527-1494
11	21.4	85.6	25	7	US-10-681-773-12160
12	20.8	83.2	481	3	US-09-918-995-11146
13	20.8	83.2	1861	3	US-09-925-065A-48863
14	20.8	83.2	5203	6	US-10-311-455-892
15	20.8	83.2	37487	7	US-10-394-948-32
16	20.8	83.2	37487	7	US-10-052-482-4
17	20.8	83.2	3673778	6	US-10-312-841-2
18	20.4	81.6	25	7	US-10-681-773-18420
19	20.2	80.8	600	9	US-10-972-079-7017
20	20.2	80.8	649	7	US-10-437-963-90579
21	20.2	80.8	5317	6	US-10-311-455-582
22	20.2	80.8	5430	7	US-10-221-714A-13
23	20.2	80.8	6880	7	US-10-221-613-183

24	20.2	80.8	9369	10	US-11-097-143-24211	Sequence 24211, A
25	20.2	80.8	10480	6	US-10-311-455-2173	Sequence 2173, Ap
26	20.2	80.8	12366	10	US-11-097-143-2608	Sequence 2608, Ap
27	19.8	79.2	490	5	US-10-027-632-6975	Sequence 6975, Ap
28	19.8	79.2	490	5	US-10-027-632-323056	Sequence 323056, Ap
29	19.8	79.2	490	6	US-10-027-632-6975	Sequence 6975, Ap
30	19.8	79.2	490	6	US-10-027-632-323056	Sequence 323056, Ap
31	19.8	79.2	849	7	US-10-437-963-91719	Sequence 91719, A
32	19.8	79.2	5919	6	US-10-311-455-161	Sequence 361, App
33	19.8	79.2	5919	7	US-10-221-613-332	Sequence 63, App
34	19.8	79.2	6181	7	US-10-221-613-332	Sequence 332, App
35	19.8	79.2	6676	7	US-10-433-793-44	Sequence 44, App
36	19.8	79.2	10254	7	US-10-433-793-44	Sequence 76, App
37	19.8	79.2	122937	7	US-10-322-281-694	Sequence 257949, A
38	19.4	77.6	616	6	US-10-027-632-257949	Sequence 257949, A
39	19.4	77.6	961	4	US-09-925-065A-55522	Sequence 55522, A
40	19.4	77.6	2000	3	US-09-938-842A-5062	Sequence 5062, Ap
41	19.4	77.6	2000	3	US-09-938-842A-5062	Sequence 5062, Ap
42	19.4	77.6	3555	5	US-10-027-632-76197	Sequence 76197, A
43	19.4	77.6	3555	6	US-10-027-632-76197	Sequence 76197, A
44	19.4	77.6	75815	8	US-10-856-218A-36	Sequence 36, App
45	19.4	77.6	75815	10	US-11-047-184-36	Sequence 36, App
46	19.4	77.6	77872	8	US-10-856-218A-44	Sequence 44, App
47	19.4	77.6	77872	10	US-11-047-184-44	Sequence 44, App
48	19.2	76.8	151	5	US-10-106-698-4136	Sequence 4308, Ap
49	19.2	76.8	360	4	US-09-983-965-4308	Sequence 119773, A
50	19.2	76.8	360	4	US-09-925-065A-119773	Sequence 119773, A
51	19.2	76.8	360	4	US-09-925-065A-119773	Sequence 119773, A
52	19.2	76.8	412	8	US-10-425-115-13144	Sequence 13144, A
53	19.2	76.8	412	8	US-10-425-115-13144	Sequence 90486, A
54	19.2	76.8	511	4	US-09-925-065A-90486	Sequence 631736, A
55	19.2	76.8	523	5	US-10-027-632-324318	Sequence 324318, A
56	19.2	76.8	523	5	US-10-027-632-324318	Sequence 324318, A
57	19.2	76.8	523	5	US-10-027-632-324318	Sequence 324318, A
58	19.2	76.8	523	6	US-10-027-632-324319	Sequence 324319, A
59	19.2	76.8	523	6	US-10-027-632-324319	Sequence 324319, A
60	19.2	76.8	540	4	US-09-925-065A-208369	Sequence 208369, A
61	19.2	76.8	541	3	US-09-864-761-15285	Sequence 15285, A
62	19.2	76.8	584	8	US-10-363-345A-26455	Sequence 26455, A
63	19.2	76.8	584	8	US-10-363-345A-26455	Sequence 26455, A
64	19.2	76.8	584	9	US-10-363-345A-26455	Sequence 26455, A
65	19.2	76.8	584	9	US-10-363-345A-26455	Sequence 26455, A
66	19.2	76.8	603	4	US-09-925-065A-658519	Sequence 658519, A
67	19.2	76.8	613	4	US-09-925-065A-261526	Sequence 261526, A
68	19.2	76.8	631	5	US-10-027-632-57738	Sequence 57738, A
69	19.2	76.8	631	5	US-10-027-632-59431	Sequence 59431, A
70	19.2	76.8	631	5	US-10-027-632-309256	Sequence 309256, A
71	19.2	76.8	631	5	US-10-027-632-309256	Sequence 309256, A
72	19.2	76.8	631	6	US-10-027-632-309257	Sequence 309257, A
73	19.2	76.8	631	6	US-10-027-632-59431	Sequence 59431, A
74	19.2	76.8	631	6	US-10-027-632-309256	Sequence 309256, A
75	19.2	76.8	631	6	US-10-027-632-309256	Sequence 309256, A
76	19.2	76.8	641	5	US-10-027-632-57738	Sequence 57738, A
77	19.2	76.8	641	5	US-10-027-632-57738	Sequence 57738, A
78	19.2	76.8	908	8	US-10-425-115-32670	Sequence 32670, A
79	19.2	76.8	1072	4	US-09-925-065A-789099	Sequence 789099, A
80	19.2	76.8	1274	4	US-09-925-065A-78982	Sequence 78982, A
81	19.2	76.8	1414	7	US-10-424-599-52605	Sequence 52605, A
82	19.2	76.8	1620	3	US-09-887-527-32	Sequence 32, App
83	19.2	76.8	1620	3	US-10-796-174-32	Sequence 32, App
84	19.2	76.8	2459	8	US-10-425-115-1578094	Sequence 1578094, A
85	19.2	76.8	2501	8	US-10-473-126-17	Sequence 17, App
86	19.2	76.8	2501	8	US-10-473-126-323	Sequence 323, App
87	19.2	76.8	2767	6	US-10-301-633-323	Sequence 23, App
88	19.2	76.8	2899	6	US-10-369-493-25859	Sequence 25859, A
89	19.2	76.8	5376	7	US-10-311-455-2124	Sequence 2124, App
90	19.2	76.8	5447	7	US-10-221-714A-484	Sequence 484, App
91	19.2	76.8	5447	6	US-10-311-455-1768	Sequence 1768, App
92	19.2	76.8	5926	6	US-10-311-455-1625	Sequence 1625, App
93	19.2	76.8	5952	7	US-10-221-613-366	Sequence 366, App
94	19.2	76.8	6030	6	US-10-311-455-153	Sequence 153, App
95	19.2	76.8	6030	5	US-10-239-676-164	Sequence 164, App
96	19.2	76.8	6030	6	US-10-240-453-186	Sequence 186, App

```
C 97 19.2 76.8 6030 7 US-10-221-613-310 Sequence 310, App
C 98 19.2 76.8 6031 7 US-10-221-714A-344 Sequence 344, App
C 99 19.2 76.8 6074 6 US-10-240-453-247 Sequence 247, App
C 100 19.2 76.8 6099 7 US-10-221-613-308 Sequence 308, App
C 101 19.2 76.8 6200 7 US-10-221-613-308 Sequence 308, App
C 102 19.2 76.8 6200 7 US-10-221-613-308 Sequence 308, App
C 103 19.2 76.8 6335 6 US-10-386-934-85 Sequence 163, App
C 104 19.2 76.8 6335 6 US-10-386-934-85 Sequence 163, App
C 105 19.2 76.8 6354 6 US-10-311-455-561 Sequence 561, App
C 106 19.2 76.8 6507 6 US-10-311-455-2007 Sequence 2007, App
C 107 19.2 76.8 7057 6 US-10-240-485-148 Sequence 148, App
C 108 19.2 76.8 7057 6 US-10-221-613-318 Sequence 318, App
C 109 19.2 76.8 7057 6 US-10-221-613-318 Sequence 318, App
C 110 19.2 76.8 7309 6 US-10-311-455-1790 Sequence 324, App
C 111 19.2 76.8 7309 6 US-10-221-613-302 Sequence 302, App
C 112 19.2 76.8 7309 6 US-10-221-613-302 Sequence 302, App
C 113 19.2 76.8 8446 6 US-10-311-455-1643 Sequence 98, App
C 114 19.2 76.8 8446 6 US-10-311-455-1643 Sequence 98, App
C 115 19.2 76.8 9515 5 US-10-239-676-159 Sequence 159, App
C 116 19.2 76.8 9515 5 US-10-239-676-159 Sequence 159, App
C 117 19.2 76.8 10552 7 US-10-221-613-111 Sequence 111, App
C 118 19.2 76.8 10552 7 US-10-221-613-111 Sequence 111, App
C 119 19.2 76.8 11260 5 US-10-239-676-119 Sequence 116, App
C 120 19.2 76.8 11260 5 US-10-239-676-119 Sequence 116, App
C 121 19.2 76.8 11296 6 US-10-311-455-696 Sequence 27, App
C 122 19.2 76.8 11577 6 US-10-311-455-696 Sequence 27, App
C 123 19.2 76.8 13792 6 US-10-311-455-1543 Sequence 416, App
C 124 19.2 76.8 13792 6 US-10-311-455-1543 Sequence 416, App
C 125 19.2 76.8 17721 6 US-10-311-455-1702 Sequence 1543, App
C 126 19.2 76.8 17918 7 US-10-221-613-381 Sequence 1702, App
C 127 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 128 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 129 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 130 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 131 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 132 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 133 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 134 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 135 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 136 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 137 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 138 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 139 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 140 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 141 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 142 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 143 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 144 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 145 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 146 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 147 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 148 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 149 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
C 150 19.2 76.8 19389 7 US-10-221-613-381 Sequence 381, App
```

ALIGNMENTS

```
RESULT 1
US-10-681-773-9
; Sequence 9, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 9
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-9
Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 1 AAAAAAAAACTTCATCATTTAAAC 25
RESULT 2
US-10-681-773-19
; Sequence 19, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 19
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-19
Query Match 96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AAAAAAAAACTTCATCATTTAAAC 25
Db 1 AAAAAAAAACTTCATCATTTAAAC 24
RESULT 3
US-10-681-773-18
; Sequence 18, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```



```
SEQ ID NO 18
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-18
```

```
Query Match          93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 44;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 1 AAAAAAAAACTTCATCATTTAAAC 25
```

```
RESULT 4
US-10-681-773-62157
Sequence 62157, Application US/10681773
Publication No. US20040146890A1
```

```
GENERAL INFORMATION:
APPLICANT: Matuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
CURRENT FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 62157
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-62157
```

```
Query Match          92.0%; Score 23; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTTCATCATTTAA 23
Db 3 AAAAAAAAACTTCATCATTTAA 25
```

```
RESULT 5
US-10-681-773-28
```

```
Sequence 28, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
CURRENT FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 28
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-28
```

```
Query Match          89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 1e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 2 AAAAAAAAACTTCATCATTTAAAC 25
Db 1 AAAAAAAAACTTCATCATTTAAAC 24
```

```
RESULT 6
US-10-681-773-18417
Sequence 18417, Application US/10681773
Publication No. US20040146890A1
```

```
GENERAL INFORMATION:
APPLICANT: Matuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
APPLICANT: Kennedy, Giulia
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
CURRENT FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 18417
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-18417
```

```
Query Match          88.0%; Score 22; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTTCATCATTTAA 22
Db 4 AAAAAAAAACTTCATCATTTAA 25
```

```
RESULT 7
US-09-920-300A-1494/C
```

```
Sequence 1494, Application US/09920300A
Patent No. US20020136728A1
GENERAL INFORMATION:
APPLICANT: King, Gordon E.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.547
CURRENT APPLICATION NUMBER: US/09/920,300A
CURRENT FILING DATE: 2001-07-31
NUMBER OF SEQ ID NOS: 1789
SOFTWARE: PatSeq for Windows Version 4.0
SEQ ID NO 1494
LENGTH: 480
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 400
OTHER INFORMATION: n = A,T,C or G
US-09-920-300A-1494
```

```
Query Match          87.2%; Score 21.8; DB 3; Length 480;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25
|||
Db 188 AAAAAAAAACTTCACATTTATAC 164

RESULT 8
US-10-033-528-1494/C
; Sequence 1494, Application US/10033528
; Publication No. US2002013197A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: SECRETIST, HEATHER
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033.528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1494
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 400
; OTHER INFORMATION: n = A,T,C or G
US-10-033-528-1494

Query Match 87.2%; Score 21.8; DB 5; Length 480;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25
|||
Db 188 AAAAAAAAACTTCACATTTATAC 164

RESULT 9
US-10-039-926-1494/C
; Sequence 1494, Application US/10099926
; Publication No. US20030166064A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secretist, Heather
; TITLE OF INVENTION: JIANG, YUJIA
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.547C2
; CURRENT APPLICATION NUMBER: US/10/099.926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1494
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 400
; OTHER INFORMATION: n = A,T,C or G
US-10-039-926-1494

Query Match 87.2%; Score 21.8; DB 6; Length 480;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25

Db 188 AAAAAAAAACTTCACATTTATAC 164
|||
|||

RESULT 10
US-10-961-527-1494/C
; Sequence 1494, Application US/10961527
; Publication No. US20050147615A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secretist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C4
; CURRENT APPLICATION NUMBER: US/10/961.527
; CURRENT FILING DATE: 2004-10-07
; PRIOR APPLICATION NUMBER: US 09/920,300
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/302,051
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/279,763
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/223,283
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1494
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 400
; OTHER INFORMATION: n = A,T,C or G
US-10-961-527-1494

Query Match 87.2%; Score 21.8; DB 9; Length 480;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25
|||
Db 188 AAAAAAAAACTTCACATTTATAC 164

RESULT 11
US-10-681-773-62160
; Sequence 62160, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methode for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681.773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 62160
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-62160

Query Match 85.6%; Score 21.4; DB 7; Length 25;

Best Local Similarity 95.7%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 23
Db 3 AAAAAAAAACTTCATCATTTAA 25

RESULT 12

US-09-918-995-11146/c
; Sequence 11146, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11146
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(481)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-11146

Query Match 83.2%; Score 20.8; DB 3; Length 481;
Best Local Similarity 91.7%; Pred. No. 6.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 458 AAAAAAAAACTTCATCATTTAA 435

RESULT 13

US-09-925-065A-48863/c
; Sequence 48863, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48863
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-48863

Query Match 83.2%; Score 20.8; DB 4; Length 1861;
Best Local Similarity 91.7%; Pred. No. 7.5e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 1042 AAAAAAAAACTTCATCATTTAA 1019

RESULT 14

US-10-311-455-892/c
; Sequence 892, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 892
; LENGTH: 5203
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-892

Query Match 83.2%; Score 20.8; DB 6; Length 5203;
Best Local Similarity 91.7%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTTCATCATTTAA 25
Db 2966 AAAAAAAAACTTCATCATTTAA 2943

RESULT 15

US-10-394-948-22/c
; Sequence 22, Application US/10394948
; Publication No. US20040023267A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: No. US20040023267A1 Compositions and Methods in Cancer
; FILE REFERENCE: 529452000900
; CURRENT APPLICATION NUMBER: US/10/394,948
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: US 60/367,025
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 37487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-394-948-22

Query Match 83.2%; Score 20.8; DB 7; Length 37487;
Best Local Similarity 91.7%; Pred. No. 1.2e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 36890 AAAAAAAAACTTCATCATTTAA 36867

```
RESULT 16
US-10-052-482-4/c
; Sequence 4, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morriss, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/799,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 37487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-4

Query Match      83.2%; Score 20.8; DB 7; Length 37487;
Best Local Similarity 91.7%; Pred. No. 1.2e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTTAA 24
Db 36890 AAAAAACATCCTTCATCATTTTAA 36867

RESULT 17
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match      83.2%; Score 20.8; DB 6; Length 3673778;
Best Local Similarity 91.7%; Pred. No. 1.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAACCTTCATCATTTTAA 25
Db 2899369 AAAAAATACCTTCATCATTTTAA 2899346

RESULT 18
US-10-681-773-18420
; Sequence 18420, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans

FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 18420
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-18420

Query Match      81.6%; Score 20.4; DB 7; Length 25;
Best Local Similarity 95.5%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTTA 22
Db 4 AAAAAAAATCTTCATCATTTTA 25

RESULT 19
US-10-972-079-7017
; Sequence 7017, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENIS, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7017
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894191582_1
US-10-972-079-7017

Query Match      80.8%; Score 20.2; DB 9; Length 600;
Best Local Similarity 88.0%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTTAA 25
Db 426 AAAAAAAGCTTCATCATTTTAA 450

RESULT 20
US-10-437-963-90579/c
; Sequence 90579, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437.963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 90579
; LENGTH: 649
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_89235C.1
US-10-437-963-90579

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 7; Length 649;
Best Local Similarity 88.0%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTTAAAC 25
Db 192 AAAAAAAAACTTCATCTTTAAAC 168

RESULT 21
US-10-311-455-582/c
; Sequence 582, Application US/10311455
; Publication No. US20030143606a1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 582
; LENGTH: 5317
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-582

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 6; Length 5317;
Best Local Similarity 88.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTTAAAC 25
Db 2161 AAAAAAAAACTTCATCTTTAAAC 2137

RESULT 22
US-10-221-714A-13/c
; Sequence 13, Application US/10221714A
; Publication No. US20040048254a1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Tumor Suppressor Genes and Oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221.714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
US-10-221-714A-13
```

```

; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 13
; LENGTH: 5430
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-13

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 7; Length 5430;
Best Local Similarity 88.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTTAAAC 25
Db 4655 AAAAAAAAACTTCATCTTTAAAC 4631

RESULT 23
US-10-221-613-183/c
; Sequence 183, Application US/10221613
; Publication No. US20040029123a1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle Regulation
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221.613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; PRIOR FILING DATE: 2001-03-15
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.00
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 183
; LENGTH: 6880
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-183

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 7; Length 6880;
Best Local Similarity 88.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTTAAAC 25
Db 796 AAAAAAAAACTTCATCTTTAAAC 772

RESULT 24
US-11-097-143-24211
```

```
; Sequence 24211, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24211
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: DROSOPHILA
; US-11-097-143-24211
```

```
Query Match      80.8%; Score 20.2; DB 10; Length 9369;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 321 AAAAAAAAAAGATTCAACATTTAAAC 345
```

```
RESULT 25
US-10-311-455-2173/c
; Sequence 2173, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2173
; LENGTH: 10480
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2173
Query Match      80.8%; Score 20.2; DB 6; Length 10480;
```

```
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 7716 AAAAAAAAACTTCATCATTTAAAC 7692
```

```
RESULT 26
US-11-097-143-2608
; Sequence 2608, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2608
; LENGTH: 12366
; TYPE: DNA
; ORGANISM: DROSOPHILA
; US-11-097-143-2608
```

```
Query Match      80.8%; Score 20.2; DB 10; Length 12366;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 321 AAAAAAAAAAGATTCAACATTTAAAC 345
```

```
RESULT 27
US-10-027-632-6975/c
; Sequence 6975, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
```


Db 72 AAAAAAAAACTTCATTGA 94

RESULT 31

US-10-437-963-91719
; Sequence 91719, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 91719
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_90267C.1
US-10-437-963-91719

Query Match 79.2%; Score 19.8; DB 7; Length 849;
Best Local Similarity 91.3%; Pred. No. 1.5e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATTGA 23

Db 445 AAAAAAAAACTTCATTGA 467

RESULT 32

US-10-311-455-361/C
; Sequence 361, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 361
; LENGTH: 5919
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-361

Query Match 79.2%; Score 19.8; DB 6; Length 5919;
Best Local Similarity 91.3%; Pred. No. 2.1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATTGA 23

Db 4059 AAAAAAAAACTTCATTGA 4037

RESULT 33

US-10-221-613-63/C
; Sequence 63, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 63
; LENGTH: 5919
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-63

Query Match 79.2%; Score 19.8; DB 7; Length 5919;
Best Local Similarity 91.3%; Pred. No. 2.1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATTGA 23

Db 4059 AAAAAAAAACTTCATTGA 4037

RESULT 34

US-10-221-613-232/C
; Sequence 232, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 232
; LENGTH: 6181
; TYPE: DNA


```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-232

Query Match
Best Local Similarity 79.2%; Score 19.8; DB 7; Length 6181;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 23
Db 1904 AAAAAAAAACTTCATCATTTAA 1882

RESULT 35
US-10-433-793-44/c
; Sequence 44, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 44
; LENGTH: 6676
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-44

Query Match
Best Local Similarity 79.2%; Score 19.8; DB 7; Length 6676;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTTCATCATTTAA 24
Db 5972 AAAAAAAAAATATCATCATTTAA 5950

RESULT 36
US-10-433-793-76/c
; Sequence 76, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 76
; LENGTH: 10254
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-76

Query Match
Best Local Similarity 79.2%; Score 19.8; DB 7; Length 10254;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 23
Db 10104 AAAAAAAAACTTCATCATTTAA 10082

RESULT 37
US-10-322-281-694
; Sequence 694, Application US/10322281
```

```
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 694
; LENGTH: 122937
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-322-281-694

Query Match
Best Local Similarity 79.2%; Score 19.8; DB 7; Length 122937;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTTCATCATTTAA 24
Db 29089 AAAAAAAAACTTCATCATTTAA 29111

RESULT 38
US-10-027-632-257949/c
; Sequence 257949, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257949
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-257949

Query Match
Best Local Similarity 77.6%; Score 19.4; DB 5; Length 616;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTTCATCATTTAA 24
Db 296 AAAAAAAAACTTCATCATTTAA 274

RESULT 39
US-10-027-632-257949/c
; Sequence 257949, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 257949
/ LENGTH: 616
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-257949
```

```
Query Match          77.6%; Score 19.4; DB 6; Length 616;
Best Local Similarity 87.0%; Pred. No. 2.1e+03;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 AAAAAAAAACTTCATCATTTAA 24
Db 296 AAAAAAAAACTTCATCATTTAA 274
```

```
RESULT 40
US-09-925-065A-56522/c
/ Sequence 56522, Application US/09925065A
/ Publication No. US20050228172A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ FILE REFERENCE: 108827.135
/ CURRENT APPLICATION NUMBER: US/09/925,065A
/ CURRENT FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243,096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289,846
/ PRIOR FILING DATE: 2001-05-09
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 56522
/ LENGTH: 961
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-925-065A-56522
```

```
Query Match          77.6%; Score 19.4; DB 4; Length 961;
Best Local Similarity 95.2%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTTCATCATTT 21
Db 270 AAAAAAAAACTTCATCATTT 250
```

```
RESULT 41
US-09-938-842A-5062
/ Sequence 5062, Application US/09938842A
/ Patent No. US20020160378A1
/ GENERAL INFORMATION:
/ APPLICANT: Harper, Jeff
/ APPLICANT: Kreps, Joel
/ APPLICANT: Wang, Xun
/ APPLICANT: Zhu, Tong
/ TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
/ FILE REFERENCE: SRIPI300-3
/ CURRENT APPLICATION NUMBER: US/09/938,842A
/ CURRENT FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 60/227,866
/ PRIOR FILING DATE: 2000-08-24
/ PRIOR APPLICATION NUMBER: US 60/264,647
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/300,111
/ PRIOR FILING DATE: 2001-06-22
/ NUMBER OF SEQ ID NOS: 5379
/ SEQ ID NO 5062
/ LENGTH: 2000
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-09-938-842A-5062
```

```
Query Match          77.6%; Score 19.4; DB 3; Length 2000;
Best Local Similarity 95.2%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTTCATCATTT 21
Db 1051 AAAAAAAAACTTCATCATTT 1071
```

```
RESULT 42
US-09-938-842A-5062
/ Sequence 5062, Application US/09938842A
/ Publication No. US20040009476A9
/ GENERAL INFORMATION:
/ APPLICANT: Harper, Jeff
/ APPLICANT: Kreps, Joel
/ APPLICANT: Wang, Xun
/ APPLICANT: Zhu, Tong
/ TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
/ FILE REFERENCE: SRIPI300-3
/ CURRENT APPLICATION NUMBER: US/09/938,842A
/ CURRENT FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 60/227,866
/ PRIOR FILING DATE: 2000-08-24
/ PRIOR APPLICATION NUMBER: US 60/264,647
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/300,111
/ PRIOR FILING DATE: 2001-06-22
/ NUMBER OF SEQ ID NOS: 5379
/ SEQ ID NO 5062
/ LENGTH: 2000
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-09-938-842A-5062
```

```
Query Match          77.6%; Score 19.4; DB 3; Length 2000;
Best Local Similarity 95.2%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTTCATCATTT 21
Db 1051 AAAAAAAAACTTCATCATTT 1071
```

```
RESULT 43
US-10-027-632-76197
; Sequence 76197, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76197
; LENGTH: 3555
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(3555)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-76197

Query Match          77.6%; Score 19.4; DB 5; Length 3555;
Best Local Similarity 87.0%; Pred. No. 2.7e+03;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 AAAAAAACCCTTCATCATTTAAAC 25
Db      1533 AAGAAGAACTCGMTCATTTAAAC 1555

RESULT 44
US-10-027-632-76197
; Sequence 76197, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 76197
; LENGTH: 3555
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(3555)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-76197

Query Match          77.6%; Score 19.4; DB 6; Length 3555;
Best Local Similarity 87.0%; Pred. No. 2.7e+03;
Matches 20; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 AAAAAAACCCTTCATCATTTAAAC 25
Db      1533 AAGAAGAACTCGMTCATTTAAAC 1555

RESULT 45
US-10-856-218A-36
; Sequence 36, Application US/10856218A
; Publication No. US20050003414A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc.
; TITLE OF INVENTION: Ovarian Cancer Promoter and Methods of Use
; FILE REFERENCE: AVI-019CIP2
; CURRENT APPLICATION NUMBER: US/10/856,218A
; CURRENT FILING DATE: 2004-05-28
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 75815
; TYPE: DNA
; ORGANISM: chicken
US-10-856-218A-36

Query Match          77.6%; Score 19.4; DB 8; Length 75815;
Best Local Similarity 95.2%; Pred. No. 4.2e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAACCCTTCATCATTT 21
Db      45769 AAAAAAACCCTTCAGCATTT 45789

RESULT 46
US-11-047-184-36
; Sequence 36, Application US/11047184
; Publication No. US20050176047A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc.
; TITLE OF INVENTION: Avian Gene Expression Controlling Regions
; FILE REFERENCE: AVI-019CIP3
; CURRENT APPLICATION NUMBER: US/11/047,184
; CURRENT FILING DATE: 2005-01-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 75815
; TYPE: DNA
; ORGANISM: chicken
US-11-047-184-36

Query Match          77.6%; Score 19.4; DB 10; Length 75815;
Best Local Similarity 95.2%; Pred. No. 4.2e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAACCCTTCATCATTT 21
Db      45769 AAAAAAACCCTTCAGCATTT 45789
```

```
RESULT 47
US-10-856-218A-44
; Sequence 44, Application US/10856218A
; Publication No. US2005000414A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc.
; TITLE OF INVENTION: Ovomucoid Promoter and Methods of Use
; FILE REFERENCE: AVI-019CIP2
; CURRENT APPLICATION NUMBER: US/10/856,218A
; CURRENT FILING DATE: 2004-05-28
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 77872
; TYPE: DNA
; ORGANISM: chicken
US-10-856-218A-44

Query Match          77.6%; Score 19.4; DB 8; Length 77872;
Best Local Similarity 95.2%; Pred. No. 4.2e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTT 21
Db 4142 AAAAAAAAACTTCAGCATTT 4162

RESULT 48
US-11-047-184-44
; Sequence 44, Application US/11047184
; Publication No. US20050176047A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc.
; TITLE OF INVENTION: Avian Gene Expression Controlling Regions
; FILE REFERENCE: AVI-019CIP3
; CURRENT APPLICATION NUMBER: US/11/047,184
; CURRENT FILING DATE: 2005-01-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 77872
; TYPE: DNA
; ORGANISM: chicken
US-11-047-184-44

Query Match          77.6%; Score 19.4; DB 10; Length 77872;
Best Local Similarity 95.2%; Pred. No. 4.2e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTT 21
Db 4142 AAAAAAAAACTTCAGCATTT 4162

RESULT 49
US-10-106-698-4136/C
; Sequence 4136, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4136

; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34)..(34)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (47)..(47)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (63)..(63)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (66)..(66)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (108)..(108)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (136)..(136)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (140)..(140)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-4136

Query Match          76.8%; Score 19.2; DB 5; Length 151;
Best Local Similarity 87.5%; Pred. No. 2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
Db 103 AAAAAAAAACTTTAATTAA 80

RESULT 50
US-09-983-965-4308
; Sequence 4308, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalegan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4308
; LENGTH: 247
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 63-LIB3058-046-Q1-K1-H4
US-09-983-965-4308

Query Match          76.8%; Score 19.2; DB 3; Length 247;
Best Local Similarity 87.5%; Pred. No. 2.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
Db 115 AAAAAAAAAACGACATCATTTATA 138

Search completed: December 14, 2005, 08:46:49
```

Thu Dec 15 09:19:19 2005

us-10-681-773-9.rnpbm

Page 15

Job time : 382.2 secs

001

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:03:56 | Search time 373.2 Seconds
(without alignments)
553.951 Million cell updates/sec

Title: US-10-681-773-10

Perfect score: 25
1 aaaaaaaccaagctgattccttc 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Published Applications NA Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	7	US-10-681-773-10
2	24.6	98.4	501	5	US-10-027-632-39077
3	24.6	98.4	501	5	US-10-027-632-39077
4	24.6	98.4	501	5	US-10-027-632-39077
5	24.6	98.4	501	5	US-10-027-632-39077
6	24.6	98.4	501	6	US-10-027-632-39077
7	24.6	98.4	501	6	US-10-027-632-39077
8	24.6	98.4	501	6	US-10-027-632-39077
9	24.6	98.4	501	6	US-10-027-632-39077
10	24.6	98.4	525	5	US-10-027-632-63078
11	24.6	98.4	525	5	US-10-027-632-63078
12	24.6	98.4	525	5	US-10-027-632-63078
13	24.6	98.4	525	5	US-10-027-632-63078
14	24.6	98.4	525	5	US-10-027-632-63078
15	24.6	98.4	525	6	US-10-027-632-63078
16	24.6	98.4	525	6	US-10-027-632-63078
17	24.6	98.4	525	6	US-10-027-632-63078
18	24.6	98.4	525	6	US-10-027-632-63078
19	24.6	98.4	525	7	US-10-681-773-20
20	24.6	98.4	525	7	US-10-681-773-20
21	24.6	98.4	525	7	US-10-681-773-20
22	24.6	98.4	525	7	US-10-681-773-20
23	24.6	98.4	525	7	US-10-681-773-20

24	22.4	89.6	1666	7	US-10-425-114-14574	Sequence 14574, A
25	21.4	85.6	25	7	US-10-681-773-87005	Sequence 87005, A
26	21.4	85.6	628	4	US-09-925-065A-800109	Sequence 800109, A
27	21.4	85.6	628	4	US-09-925-065A-800110	Sequence 800110, A
28	21.4	85.6	628	4	US-09-925-065A-800111	Sequence 800111, A
29	21.4	85.6	628	4	US-09-925-065A-854017	Sequence 854017, A
30	21.4	85.6	25	7	US-10-681-773-251	Sequence 251, App
31	20.2	80.8	317	8	US-10-425-115-67105	Sequence 67105, A
32	20.2	80.8	439	4	US-09-925-065A-639643	Sequence 639643, A
33	20.2	80.8	548	4	US-09-925-065A-79960	Sequence 79960, A
34	20.2	80.8	586	4	US-09-925-065A-637117	Sequence 637117, A
35	20.2	80.8	591	4	US-09-925-065A-725331	Sequence 725331, A
36	20.2	80.8	606	4	US-09-925-065A-743446	Sequence 743446, A
37	20.2	80.8	807	5	US-10-027-632-154034	Sequence 154034, A
38	20.2	80.8	4254	10	US-11-097-143-4504	Sequence 4504, Ap
39	20.2	80.8	6746	10	US-11-097-143-2875	Sequence 2875, Ap
40	20.2	80.8	61020	7	US-10-221-714A-514	Sequence 514, App
41	20.2	80.8	235070	5	US-10-087-192-1990	Sequence 1990, Ap
42	20.2	80.8	521	4	US-09-925-065A-334939	Sequence 334939, A
43	19.8	79.2	13606	5	US-10-239-676-166	Sequence 166, App
44	19.8	79.2	13606	6	US-10-311-455-1784	Sequence 1784, Ap
45	19.8	79.2	13606	6	US-10-240-453-188	Sequence 188, App
46	19.8	79.2	13606	6	US-10-221-714A-284	Sequence 284, App
47	19.8	79.2	13606	7	US-10-681-773-258	Sequence 258, App
48	19.4	77.6	201	8	US-10-719-993-33249	Sequence 33249, A
49	19.2	76.8	201	8	US-10-357-930-7904	Sequence 7904, Ap
50	19.2	76.8	387	8	US-10-357-930-7442	Sequence 7442, Ap
51	19.2	76.8	400	8	US-10-357-930-4072	Sequence 4072, Ap
52	19.2	76.8	424	8	US-09-925-065A-666159	Sequence 666159, A
53	19.2	76.8	503	4	US-09-925-065A-666160	Sequence 666160, A
54	19.2	76.8	532	4	US-09-925-065A-594044	Sequence 594044, A
55	19.2	76.8	537	4	US-09-925-065A-40292	Sequence 40292, A
56	19.2	76.8	537	4	US-09-925-065A-159837	Sequence 159837, A
57	19.2	76.8	567	4	US-09-925-065A-600459	Sequence 600459, A
58	19.2	76.8	567	4	US-09-925-065A-600460	Sequence 600460, A
59	19.2	76.8	567	4	US-09-925-065A-600461	Sequence 600461, A
60	19.2	76.8	567	4	US-10-027-632-303508	Sequence 303508, A
61	19.2	76.8	601	5	US-10-027-632-303508	Sequence 303508, A
62	19.2	76.8	601	5	US-10-424-559A-33657	Sequence 33657, A
63	19.2	76.8	765	7	US-09-925-065A-155472	Sequence 155472, A
64	19.2	76.8	1089	4	US-10-425-115-101300	Sequence 101300, A
65	19.2	76.8	1335	8	US-10-425-115-101300	Sequence 101300, A
66	19.2	76.8	1834	6	US-11-013-031A-105	Sequence 105, App
67	19.2	76.8	1834	6	US-11-012-546-105	Sequence 105, App
68	19.2	76.8	1834	6	US-09-925-065A-13347	Sequence 13347, A
69	19.2	76.8	1889	4	US-10-481-032A-760	Sequence 760, App
70	19.2	76.8	2000	7	US-10-260-238-2650	Sequence 2650, Ap
71	19.2	76.8	2031	8	US-10-425-115-79256	Sequence 79256, A
72	19.2	76.8	6048	6	US-10-311-455-482	Sequence 482, App
73	19.2	76.8	6051	8	US-10-723-860-3834	Sequence 3834, App
74	19.2	76.8	8333	5	US-10-239-676-113	Sequence 113, App
75	19.2	76.8	8333	5	US-10-311-455-1175	Sequence 1175, App
76	19.2	76.8	8333	6	US-10-240-453-119	Sequence 119, App
77	19.2	76.8	8333	6	US-10-311-455-119	Sequence 119, App
78	19.2	76.8	8333	6	US-10-311-455-119	Sequence 119, App
79	19.2	76.8	8333	6	US-10-311-455-119	Sequence 119, App
80	19.2	76.8	9483	7	US-10-221-613-48	Sequence 48, App
81	19.2	76.8	13996	5	US-09-764-870-602	Sequence 602, App
82	19.2	76.8	13996	5	US-10-125-540-602	Sequence 602, App
83	19.2	76.8	14001	3	US-09-764-870-601	Sequence 601, App
84	19.2	76.8	14001	3	US-10-125-540-601	Sequence 601, App
85	19.2	76.8	26230	5	US-10-087-192-1246	Sequence 1246, Ap
86	19.2	76.8	28433	10	US-11-097-143-11950	Sequence 11950, A
87	19.2	76.8	34433	7	US-10-052-482-116	Sequence 116, App
88	19.2	76.8	45027	8	US-10-417-375-58	Sequence 58, App
89	19.2	76.8	96597	3	US-09-997-722-73	Sequence 73, App
90	19.2	76.8	105730	3	US-10-741-600-17809	Sequence 17809, A
91	19.2	76.8	168974	6	US-10-085-117-139	Sequence 139, App
92	19.2	76.8	187844	6	US-10-719-993-6883	Sequence 6883, Ap
93	19.2	76.8	196686	5	US-10-087-192-484	Sequence 484, App
94	19.2	76.8	197775	5	US-10-087-192-853	Sequence 853, App
95	19.2	76.8	659158	3	US-09-771-208-20	Sequence 20, App
96	18.8	75.2	43	3	US-09-932-165-1485	Sequence 1485, Ap

```

C 97 18.8 75.2 43 3 US-09-942-052-714 Sequence 714, App
C 98 18.8 75.2 43 5 US-10-121-019-13 Sequence 13, Appl
C 99 18.8 75.2 43 6 US-10-013-312-2295 Sequence 2295, Ap
C 100 18.8 75.2 43 6 US-10-087-190-30 Sequence 30, Appl
C 101 18.8 75.2 43 6 US-10-120-885A-28 Sequence 28, Appl
C 102 18.8 75.2 43 6 US-10-121-016-55 Sequence 55, Appl
C 103 18.8 75.2 43 7 US-10-114-669-6 Sequence 6, Appl
C 104 18.8 75.2 43 7 US-10-114-432-42 Sequence 42, Appl
C 105 18.8 75.2 44 3 US-09-771-312-7 Sequence 7, Appl
C 106 18.8 75.2 124 5 US-10-121-019-4 Sequence 4, Appl
C 107 18.8 75.2 192 9 US-10-837-269-1 Sequence 1, Appl
C 108 18.8 75.2 346 5 US-10-121-019-1 Sequence 1, Appl
C 109 18.8 75.2 425 3 US-09-771-312-3 Sequence 3, Appl
C 110 18.8 75.2 560 4 US-09-925-065A-149624 Sequence 149624,
C 111 18.8 75.2 560 4 US-09-925-065A-149625 Sequence 149625,
C 112 18.8 75.2 560 4 US-09-925-065A-149626 Sequence 149626,
C 113 18.8 75.2 581 4 US-09-925-065A-937904 Sequence 937904,
C 114 18.8 75.2 581 4 US-09-925-065A-937905 Sequence 937905,
C 115 18.8 75.2 606 4 US-09-925-065A-737130 Sequence 737130,
C 116 18.8 75.2 606 4 US-09-925-065A-814497 Sequence 814497,
C 117 18.8 75.2 686 8 US-10-363-345A-16389 Sequence 16389, A
C 118 18.8 75.2 686 8 US-10-363-345A-16390 Sequence 16390, A
C 119 18.8 75.2 686 9 US-10-363-483A-16389 Sequence 16389, A
C 120 18.8 75.2 686 9 US-10-363-483A-16390 Sequence 16390, A
C 121 18.8 75.2 2000 3 US-09-938-842A-2750 Sequence 2750, Ap
C 122 18.8 75.2 2000 3 US-10-102-524-1482 Sequence 2750, Ap
C 123 18.6 74.4 115 5 US-10-242-535A-12962 Sequence 12962, A
C 124 18.6 74.4 140 7 US-10-085-783A-12962 Sequence 12962, A
C 125 18.6 74.4 140 7 US-10-950-009-309 Sequence 309, App
C 126 18.6 74.4 167 9 US-10-425-115-55214 Sequence 55214, A
C 127 18.6 74.4 172 8 US-10-741-601-20661 Sequence 20661, A
C 128 18.6 74.4 201 7 US-10-719-993-11473 Sequence 11473, A
C 129 18.6 74.4 201 8 US-10-425-115-105545 Sequence 105545,
C 130 18.6 74.4 255 8 US-10-425-115-146440 Sequence 146440,
C 131 18.6 74.4 307 8 US-10-425-115-142042 Sequence 142042,
C 132 18.6 74.4 308 8 US-10-425-115-94846 Sequence 94846, A
C 133 18.6 74.4 331 9 US-10-425-115-85675 Sequence 85675, A
C 134 18.6 74.4 344 8 US-09-925-065A-782593 Sequence 782593,
C 135 18.6 74.4 370 4 US-09-920-300A-480 Sequence 480, App
C 136 18.6 74.4 381 5 US-10-033-528-480 Sequence 480, App
C 137 18.6 74.4 381 5 US-10-099-925-480 Sequence 480, App
C 138 18.6 74.4 381 5 US-10-961-527-480 Sequence 124757,
C 139 18.6 74.4 390 8 US-10-425-115-124757 Sequence 124757,
C 140 18.6 74.4 408 7 US-10-437-963-18963 Sequence 18963, A
C 141 18.6 74.4 446 5 US-10-027-632-43399 Sequence 43399, A
C 142 18.6 74.4 446 5 US-10-027-632-43399 Sequence 43399, A
C 143 18.6 74.4 482 5 US-10-027-632-72395 Sequence 72395, A
C 144 18.6 74.4 482 5 US-10-027-632-72395 Sequence 72395, A
C 145 18.6 74.4 485 5 US-10-027-632-312707 Sequence 312707,
C 146 18.6 74.4 485 6 US-10-027-632-312707 Sequence 312707,
C 147 18.6 74.4 495 8 US-10-425-115-50194 Sequence 50194, A
C 148 18.6 74.4 495 8 US-10-425-115-50194 Sequence 50194, A
C 149 18.6 74.4 495 8 US-10-425-115-50194 Sequence 50194, A
C 150 18.6 74.4 495 8 US-10-425-115-50194 Sequence 50194, A
```

ALIGNMENTS

```

RESULT 1
US-10-681-773-10
; Sequence 10, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
```

```

; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 10
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-10
```

```

Query Match 100.0%; Score 25; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAACTAAAGCTTGATCTTC 25
Db 1 AAAAAAAAACTAAAGCTTGATCTTC 25
```

```

RESULT 2
US-10-027-632-39077
; Sequence 39077, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-05
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39077
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39077

Query Match 98.4%; Score 24.6; DB 5; Length 501;
Best Local Similarity 96.0%; Pred. No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAACTAAAGCTTGATCTTC 25
Db 347 AAAAAAAAACTAAAGCTTGATCTTC 371
```

```

RESULT 3
US-10-027-632-39078
; Sequence 39078, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```



```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39078
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39078
```

```

Query Match          98.4%: Score 24.6; DB 5; Length 501;
Best Local Similarity 96.0%: Pred No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
Db      347 AAAAAAAAACTAAGCTTGATCTTC 371
```

```

RESULT 4
US-10-027-632-39079
; Sequence 39079, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39079
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39079
```

```

Query Match          98.4%: Score 24.6; DB 5; Length 501;
Best Local Similarity 96.0%: Pred No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
Db      347 AAAAAAAAACTAAGCTTGATCTTC 371
```

```

RESULT 5
US-10-027-632-39080
; Sequence 39080, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39080
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-39080
```

```

Query Match          98.4%: Score 24.6; DB 5; Length 501;
Best Local Similarity 96.0%: Pred No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
Db      347 AAAAAAAAACTAAGCTTGATCTTC 371
```

```

RESULT 6
US-10-027-632-39077
; Sequence 39077, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39077
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-39077

Query Match 98.4%; Score 24.6; DB 6; Length 501;
Best Local Similarity 96.0%; Pred. No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||:|||||
Db 347 AAAAAAAAACTAAGCTTGATCTTC 371

RESULT 7
US-10-027-632-39078

Sequence 39078, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO: 39078

LENGTH: 501

TYPE: DNA

ORGANISM: Human

US-10-027-632-39078

Query Match 98.4%; Score 24.6; DB 6; Length 501;
Best Local Similarity 96.0%; Pred. No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||:|||||
Db 347 AAAAAAAAACTAAGCTTGATCTTC 371

RESULT 8
US-10-027-632-39079

Sequence 39079, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO: 39079

LENGTH: 501

TYPE: DNA

ORGANISM: Human

US-10-027-632-39079

Query Match 98.4%; Score 24.6; DB 6; Length 501;
Best Local Similarity 96.0%; Pred. No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||:|||||
Db 347 AAAAAAAAACTAAGCTTGATCTTC 371

RESULT 9
US-10-027-632-39080

Sequence 39080, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO: 39080

LENGTH: 501

TYPE: DNA

ORGANISM: Human

US-10-027-632-39080

Query Match 98.4%; Score 24.6; DB 6; Length 501;
Best Local Similarity 96.0%; Pred. No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||:|||||
Db 347 AAAAAAAAACTAAGCTTGATCTTC 371

RESULT 10
US-10-027-632-63078

Sequence 63078, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63078
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63078

Query Match
Best Local Similarity 98.4%; Score 24.6; DB 5; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 11
US-10-027-632-63079
; Sequence 63079, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63079
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63079

Query Match
Best Local Similarity 98.4%; Score 24.6; DB 5; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395
```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63080
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63080

Query Match
Best Local Similarity 98.4%; Score 24.6; DB 5; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 12
US-10-027-632-63080
; Sequence 63080, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63080
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63080

Query Match
Best Local Similarity 98.4%; Score 24.6; DB 5; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 13
US-10-027-632-63081
; Sequence 63081, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63081
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63081
```

```
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63081

Query Match
  Best Local Similarity 98.4%; Score 24.6; DB 6; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 14
US-10-027-632-63078
; Sequence 63078, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63078
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63078

Query Match
  Best Local Similarity 98.4%; Score 24.6; DB 6; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 15
US-10-027-632-63079
; Sequence 63079, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63079
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63080

Query Match
  Best Local Similarity 98.4%; Score 24.6; DB 6; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 16
US-10-027-632-63080
; Sequence 63080, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63080
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63080

Query Match
  Best Local Similarity 98.4%; Score 24.6; DB 6; Length 525;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 17
US-10-027-632-63081
; Sequence 63081, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63081
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63081

Query Match          98.4%; Score 24.6; DB 6; Length 525;
Best Local Similarity 96.0%; Pred. No. 7.1;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 371 AAAAAAAAACTAAGCTTGATCTTC 395

RESULT 18
US-10-681-773-3
; Sequence 3, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 3
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-3

Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 24
    |||||
Db 2 AAAAAAAAACTAAGCTTGATCTT 25

RESULT 19
US-10-681-773-20
; Sequence 20, Application US/10681773
; Publication No. US20040146890A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 20
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-20

Query Match          96.0%; Score 24; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 1 AAAAAAAAACTAAGCTTGATCTTC 24

RESULT 20
US-10-681-773-11
; Sequence 11, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 11
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-11

Query Match          93.6%; Score 23.4; DB 7; Length 25;
Best Local Similarity 96.0%; Pred. No. 12;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 1 AAAAAAAAACTAAGCTTGATCTTC 25

RESULT 21
US-10-681-773-87004
; Sequence 87004, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
```

```
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO: 87004
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-87004

Query Match          92.0%; Score 23; DB 7; Length 25;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 23
   |||||
Db 3 AAAAAAAAACTAAGCTTGATCT 25

RESULT 22
US-10-681-773-4
/ Sequence 4, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsuzaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO: 4
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-4

Query Match          89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 31;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 24
   |||||
Db 2 AAAAAAAAACTAAGCTTGATCTT 25

RESULT 23
US-10-681-773-21
/ Sequence 21, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsuzaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```

```
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
/ SEQ ID NO: 21
/ LENGTH: 25
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-10-681-773-21

Query Match          89.6%; Score 22.4; DB 7; Length 25;
Best Local Similarity 95.8%; Pred. No. 31;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
   |||||
Db 1 AAAAAAAAACTAAGCTTGATCTTC 24

RESULT 24
US-10-425-114-14574
/ Sequence 14574, Application US/10425114
/ Publication No. US20040034888A1
/ GENERAL INFORMATION:
/ APPLICANT: Zhou, Jingdong
/ APPLICANT: Liu, Jingdong
/ APPLICANT: Kowalic, David K.
/ APPLICANT: Screen, Steven E.
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO: 14574
/ LENGTH: 1666
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
/ FEATURE:
/ OTHER INFORMATION: Clone ID: L1B23-006-F9_FLI
US-10-425-114-14574

Query Match          89.6%; Score 22.4; DB 7; Length 1666;
Best Local Similarity 95.8%; Pred. No. 66;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
   |||||
Db 1643 AAAAAAAAACTAAGCTTGATCTTC 1666

RESULT 25
US-10-681-773-87005
/ Sequence 87005, Application US/10681773
/ Publication No. US20040146890A1
/ GENERAL INFORMATION:
/ APPLICANT: Matsuzaki, Hajime
/ APPLICANT: Mei, Rui
/ APPLICANT: Shen, Mei-Mei
/ APPLICANT: Kennedy, Giulia
/ TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
/ FILE REFERENCE: 3522.2
/ CURRENT APPLICATION NUMBER: US/10/681,773
/ CURRENT FILING DATE: 2003-10-07
/ PRIOR APPLICATION NUMBER: 60/470,475
/ PRIOR FILING DATE: 2002-05-14
/ PRIOR APPLICATION NUMBER: 60/417,190
/ PRIOR FILING DATE: 2002-10-08
/ NUMBER OF SEQ ID NOS: 124031
/ SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
```

```
; SEQ ID NO 87005
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-681-773-87005

Query Match      85.6%; Score 21.4; DB 7; Length 25;
Best Local Similarity 95.7%; Pred. No. 76;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCT 23
        |||||||
Db      3 AAAAAAAAACTAAGCTTGATCT 25

RESULT 26
US-09-925-065A-800109/c
; Sequence 800109, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800109
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-800109

Query Match      85.6%; Score 21.4; DB 4; Length 628;
Best Local Similarity 95.7%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCT 23
        |||||||
Db      280 AAAAAAAAACTAAGCTTGATCT 258

RESULT 27
US-09-925-065A-800110/c
; Sequence 800110, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
```

```
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800110
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-800110

Query Match      85.6%; Score 21.4; DB 4; Length 628;
Best Local Similarity 95.7%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCT 23
        |||||||
Db      280 AAAAAAAAACTAAGCTTGATCT 258

RESULT 28
US-09-925-065A-800111/c
; Sequence 800111, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 800111
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-800111

Query Match      85.6%; Score 21.4; DB 4; Length 628;
Best Local Similarity 95.7%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCT 23
        |||||||
Db      280 AAAAAAAAACTAAGCTTGATCT 258

RESULT 29
US-09-925-065A-854017/c
; Sequence 854017, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
```

```

; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 854017
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-854017

Query Match
Best Local Similarity 85.6%; Score 21.4; DB 4; Length 628;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 23
Db 280 AAAAAAAAAAGCTTAAGCTTGATCTT 258

RESULT 30
US-10-681-773-251
; Sequence 251, Application US/10681773
; Publication No. US20040146890A1
; GENERAL INFORMATION:
; APPLICANT: Matsuzaki, Hajime
; APPLICANT: Mei, Rui
; APPLICANT: Shen, Mei-Mei
; APPLICANT: Kennedy, Giulia
; TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
; FILE REFERENCE: 3522.2
; CURRENT APPLICATION NUMBER: US/10/681,773
; PRIOR FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/470,475
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/417,190
; PRIOR FILING DATE: 2002-10-08
; NUMBER OF SEQ ID NOS: 124031
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 251
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-681-773-251

Query Match
Best Local Similarity 84.0%; Score 21; DB 7; Length 25;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 AAAAACTAAAGCTTGATCTTC 25
Db 1 AAAAACTAAAGCTTGATCTTC 21

RESULT 31
US-10-425-115-67105
; Sequence 67105, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 67105
; LENGTH: 317
; TYPE: DNA
```

```

; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_161198C.1
US-10-425-115-67105

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 8; Length 317;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTTC 25
Db 114 AAAAAAAAAAGTAAATCTTGATCTCC 138

RESULT 32
US-09-925-065A-639643
; Sequence 639643, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 639643
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-639643

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 4; Length 439;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTTC 25
Db 255 AAAAAATAAAGTAAGCTTGATCTTC 279

RESULT 33
US-09-925-065A-79960
; Sequence 79960, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
```



```

; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79960
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-79960
```

```

Query Match      80.8%; Score 20.2; DB 4; Length 548;
Best Local Similarity 88.0%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
         ||||| ||||| ||||| ||||| |||||
Db      195 AAAAAAACACCAATCTTGATCTTC 219
```

RESULT 34

```

US-09-925-065A-637117/c
; Sequence 637117, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 637117
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-637117
```

```

Query Match      80.8%; Score 20.2; DB 4; Length 586;
Best Local Similarity 88.0%; Pred. No. 4.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
         ||||| ||||| ||||| ||||| |||||
Db      249 AAAATRAAGATAAGCTTGATCTTC 225
```

RESULT 35

```

US-09-925-065A-272531
; Sequence 272531, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
```

```

; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272531
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-272531
```

```

Query Match      80.8%; Score 20.2; DB 4; Length 591;
Best Local Similarity 88.0%; Pred. No. 4.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
         ||||| ||||| ||||| ||||| |||||
Db      90 AAAACAAAAGTAAGCTTAATCTTC 114
```

RESULT 36

```

US-09-925-065A-743446
; Sequence 743446, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 743446
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-743446
```

```

Query Match      80.8%; Score 20.2; DB 4; Length 606;
Best Local Similarity 88.0%; Pred. No. 4.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
         ||||| ||||| ||||| ||||| |||||
Db      19 AAAACAAAAGTAAGCTTAATCTTC 43
```

RESULT 37

```

US-10-027-632-154034/c
; Sequence 154034, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
```

```
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 154034
;; LENGTH: 807
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-154034
```

```
Query Match      80.8%; Score 20.2; DB 5; Length 807;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 294 AAAAAAACACCAATCTTGATCTTC 270
```

```
RESULT 38
US-10-027-632-154034/c
; Sequence 154034, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; FILE REFERENCE: 108827.129
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US/10/027,632
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 154034
;; LENGTH: 807
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-154034
```

```
Query Match      80.8%; Score 20.2; DB 6; Length 807;
Best Local Similarity 88.0%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 294 AAAAAAACACCAATCTTGATCTTC 270
```

```
RESULT 39
US-11-097-143-4504/c
; Sequence 4504, Application US/11097143
; Publication No. US20050208558A1
```

```
;; GENERAL INFORMATION:
;; APPLICANT: Venter, J. Craig
;; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;; FILE REFERENCE: CL000728
;; CURRENT APPLICATION NUMBER: US/11/097,143
;; CURRENT FILING DATE: 2005-04-04
;; PRIOR APPLICATION NUMBER: 60/157,832
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: 60/160,191
;; PRIOR FILING DATE: 1999-10-19
;; PRIOR APPLICATION NUMBER: 60/161,932
;; PRIOR FILING DATE: 1999-10-28
;; PRIOR APPLICATION NUMBER: 60/164,769
;; PRIOR FILING DATE: 1999-11-12
;; PRIOR APPLICATION NUMBER: 60/173,383
;; PRIOR FILING DATE: 1999-12-28
;; PRIOR APPLICATION NUMBER: 60/175,693
;; PRIOR FILING DATE: 2000-01-12
;; PRIOR APPLICATION NUMBER: 60/184,831
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: 60/191,637
;; PRIOR FILING DATE: 2000-03-23
;; NUMBER OF SEQ ID NOS: 43008
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4504
;; LENGTH: 4254
;; TYPE: DNA
;; ORGANISM: DROSOPHILA
US-11-097-143-4504
```

```
Query Match      80.8%; Score 20.2; DB 10; Length 4254;
Best Local Similarity 88.0%; Pred. No. 5.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 826 AAAAAAACCAATCTTGATCTTC 802
```

```
RESULT 40
US-11-097-143-2875/c
; Sequence 2875, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
;; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;; FILE REFERENCE: CL000728
;; CURRENT APPLICATION NUMBER: US/11/097,143
;; CURRENT FILING DATE: 2005-04-04
;; PRIOR APPLICATION NUMBER: 60/157,832
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: 60/160,191
;; PRIOR FILING DATE: 1999-10-19
;; PRIOR APPLICATION NUMBER: 60/161,932
;; PRIOR FILING DATE: 1999-10-28
;; PRIOR APPLICATION NUMBER: 60/164,769
;; PRIOR FILING DATE: 1999-11-12
;; PRIOR APPLICATION NUMBER: 60/173,383
;; PRIOR FILING DATE: 1999-12-28
;; PRIOR APPLICATION NUMBER: 60/175,693
;; PRIOR FILING DATE: 2000-01-12
;; PRIOR APPLICATION NUMBER: 60/184,831
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: 60/191,637
;; PRIOR FILING DATE: 2000-03-23
;; NUMBER OF SEQ ID NOS: 43008
;; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 2875
; LENGTH: 6746
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-2875

Query Match      80.8%; Score 20.2; DB 10; Length 6746;
Best Local Similarity 88.0%; Pred. No. 6.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCTTC 25
Db      5952 AACCAAACTAAGCTCGATATTC 5928

RESULT 41
US-10-221-714A-514/C
; Sequence 514, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: BERLIN, Kurt
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; FILE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 514
; LENGTH: 61020
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-514

Query Match      80.8%; Score 20.2; DB 7; Length 61020;
Best Local Similarity 88.0%; Pred. No. 9.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCTTC 25
Db      25658 AAAAAAACTAATTTATCTTC 25634

RESULT 42
US-10-087-192-1990/C
; Sequence 1990, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: CANCER
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
```

```
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1990
; LENGTH: 235070
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(235070)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1990

Query Match      80.8%; Score 20.2; DB 5; Length 235070;
Best Local Similarity 88.0%; Pred. No. 1.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCTTC 25
Db      126963 AAAAAAAGATGATCTTC 126939

RESULT 43
US-09-925-065A-334939/C
; Sequence 334939, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334939
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-334939

Query Match      79.2%; Score 19.8; DB 4; Length 521;
Best Local Similarity 91.3%; Pred. No. 5.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCT 23
Db      37 AAAAAAAGTTTATCT 15

RESULT 44
US-10-239-676-166/C
; Sequence 166, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
```

```
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 166
LENGTH: 13606
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (5950, 5973, 8958)
US-10-239-676-166
```

Query Match 79.2%; Score 19.8; DB 5; Length 13606;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAACTAAGCTTGATCT 23
DB 5438 AAAAAAAAACTAAGCTTAATCT 5416
```

```
RESULT 45
US-10-311-455-1784/c
Sequence 1784, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
FILE REFERENCE: 5013.1014
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1784
LENGTH: 13606
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (5950, 5973, 8958)
OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1784
```

Query Match 79.2%; Score 19.8; DB 6; Length 13606;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAACTAAGCTTGATCT 23
DB 5438 AAAAAAAAACTAAGCTTAATCT 5416
```

RESULT 46
US-10-240-453-188/c

```
Sequence 188, Application US/10240453
Publication No. US20030148326A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
TITLE OF INVENTION: Transcription
TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
FILE REFERENCE: 5013.1009
CURRENT FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: US/10/240,453
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 350
SEQ ID NO 188
LENGTH: 13606
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (5950, 5973, 8958)
US-10-240-453-188
```

Query Match 79.2%; Score 19.8; DB 6; Length 13606;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAACTAAGCTTGATCT 23
DB 5438 AAAAAAAAACTAAGCTTAATCT 5416
```

```
RESULT 47
US-10-221-714A-284/c
Sequence 284, Application US/10221714A
Publication No. US20040048254A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
TITLE OF INVENTION: tumor suppressor genes and oncogenes
FILE REFERENCE: 5013.1005
CURRENT FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: PCT/EP01/02955
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: DE 10013847.0
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 540
SEQ ID NO 284
LENGTH: 13606
TYPE: DNA
```

```
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: 5950, 5973, 8958
OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-284
```

Query Match 79.2%; Score 19.8; DB 7; Length 13606;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 23
Db 5438 AAAAAAAAACTAAGCTTGATCT 5416

```
RESULT 48
US-10-681-773-258
Sequence 258, Application US/10681773
Publication No. US20040146890A1
GENERAL INFORMATION:
APPLICANT: Matsuzaki, Hajime
APPLICANT: Mei, Rui
APPLICANT: Shen, Mei-Mei
TITLE OF INVENTION: Methods for Genotyping Polymorphisms in Humans
FILE REFERENCE: 3522.2
CURRENT APPLICATION NUMBER: US/10/681,773
CURRENT FILING DATE: 2003-10-07
PRIOR APPLICATION NUMBER: 60/470,475
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 60/417,190
PRIOR FILING DATE: 2002-10-08
NUMBER OF SEQ ID NOS: 124031
SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
SEQ ID NO 258
LENGTH: 25
TYPE: DNA
ORGANISM: Homo sapien
US-10-681-773-258
```

Query Match 77.6%; Score 19.4; DB 7; Length 25;
Best Local Similarity 95.2%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAAAACTAAGCTTGATCTTC 25
Db 1 AAAAACTAAGCTTGATCTTC 21

```
RESULT 49
US-10-719-993-39249/C
Sequence 39249, Application US/10719993
Publication No. US20040265849A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001496
CURRENT APPLICATION NUMBER: US/10/719,993
CURRENT FILING DATE: 2003-11-24
NUMBER OF SEQ ID NOS: 55342
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39249
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-719-993-39249
```

Query Match 76.8%; Score 19.2; DB 8; Length 201;
Best Local Similarity 87.5%; Pred. No. 8.3e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAACTAAGCTTGATCTT 24
Db 37 AAAAAAACTAAGCTTGATCTT 14

```
RESULT 50
US-10-357-930-7904/C
Sequence 7904, Application US/10357930
Publication No. US20040259086A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Endege, Wilson
APPLICANT: Monahan, John
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
FILE REFERENCE: MRI-007BCN
CURRENT APPLICATION NUMBER: US/10/357,930
CURRENT FILING DATE: 2003-02-04
PRIOR APPLICATION NUMBER: 09/785,276
PRIOR FILING DATE: 2003-02-16
PRIOR APPLICATION NUMBER: 60/183,319
PRIOR FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/189,862
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/207,454
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: 60/211,314
PRIOR FILING DATE: 2000-06-09
PRIOR APPLICATION NUMBER: 60/219,007
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/255,281
PRIOR FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 62232
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7904
LENGTH: 387
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 369
OTHER INFORMATION: n = A,T,C or G
US-10-357-930-7904
```

Query Match 76.8%; Score 19.2; DB 8; Length 387;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAACTAAGCTTGATCTT 24
Db 364 AAAAAAACTAAGCTTGATCTT 341

Search completed: December 14, 2005, 08:46:51
Job time : 375.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:43:33 ; Search time 180.2 Seconds
(without alignments)
68.002 Million cell updates/sec

Title: US-10-681-773-6

Perfect score: 25

Sequence: 1 aaaaaaaaaagcatgtatgtacac 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4161359 seqs, 245077644 residues

Total number of hits satisfying chosen parameters: 8322718

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications NA.New:*
1: /cgn2_6/ptodaca/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/ptodaca/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodaca/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodaca/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/ptodaca/1/pubpna/PCR_NEW_PUB.seq:*
6: /cgn2_6/ptodaca/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodaca/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/ptodaca/1/pubpna/US11_NEW_PUB.seq2:*
9: /cgn2_6/ptodaca/1/pubpna/US11_NEW_PUB.seq3:*
10: /cgn2_6/ptodaca/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.8	79.2	596	US-10-750-185-20178	Sequence 20178, A
2	19.8	79.2	875	US-10-750-185-39819	Sequence 39819, A
3	19.4	77.6	246960	US-11-121-086-8	Sequence 8, Appl1
4	19.2	76.8	737	US-10-750-185-35652	Sequence 35652, A
5	19.2	76.8	2275	US-10-750-185-61429	Sequence 61429, A
6	18.6	75.2	3571	US-10-750-185-53588	Sequence 53588, A
7	18.6	74.4	2049	US-10-750-185-40699	Sequence 40699, A
8	18.6	74.4	2477	US-10-750-185-56281	Sequence 56281, A
9	18.6	74.4	3138	US-10-750-185-45175	Sequence 45175, A
10	18.6	74.4	158632	US-11-121-086-30	Sequence 30, Appl1
11	18.6	74.4	171427	US-11-112-908-60	Sequence 60, Appl1
12	18.6	74.4	222094	US-10-995-561-13244	Sequence 13244, A
13	18.4	73.6	169495	US-11-121-086-61	Sequence 61, Appl1
14	18.2	72.8	1282	US-10-750-185-47195	Sequence 47195, A
15	18.2	72.8	131855	US-11-112-908-29	Sequence 29, Appl1
16	18.2	72.8	150173	US-11-112-908-26	Sequence 26, Appl1
17	18.2	72.8	166030	US-11-112-908-28	Sequence 28, Appl1
18	18.2	72.8	171247	US-11-112-908-27	Sequence 27, Appl1
19	18.2	72.8	218821	US-11-121-086-31	Sequence 31, Appl1
20	18	72.0	201	US-10-995-561-57202	Sequence 57202, A
21	18	72.0	98716	US-10-995-561-13331	Sequence 13331, A
22	17.8	71.2	201	US-10-995-561-83376	Sequence 83376, A
23	17.8	71.2	3734	US-10-750-185-45080	Sequence 45080, A

C 24	17.8	71.2	17797	6	US-10-995-561-13506	Sequence 13506, A
C 25	17.8	71.2	110000	7	US-11-155-492-1	Sequence 1, Appl1
C 26	17.8	71.2	180654	7	US-11-121-086-58	Sequence 58, Appl1
C 27	17.6	70.4	201	6	US-10-995-561-31582	Sequence 31582, A
C 28	17.6	70.4	201	6	US-10-995-561-31727	Sequence 31727, A
C 29	17.6	70.4	201	6	US-10-995-561-31728	Sequence 31728, A
C 30	17.6	70.4	201	6	US-10-995-561-58952	Sequence 58952, A
C 31	17.6	70.4	201	6	US-10-995-561-69364	Sequence 69364, A
C 32	17.6	70.4	600	6	US-10-750-185-618	Sequence 618, App
C 33	17.6	70.4	979	6	US-10-750-185-48111	Sequence 48111, A
C 34	17.6	70.4	1049	6	US-10-750-185-45903	Sequence 45903, A
C 35	17.6	70.4	1193	6	US-10-750-185-59829	Sequence 59829, A
C 36	17.6	70.4	1328	6	US-10-750-185-30513	Sequence 30513, A
C 37	17.6	70.4	1760	6	US-10-750-185-64253	Sequence 64253, A
C 38	17.6	70.4	2051	6	US-10-750-185-64736	Sequence 64736, A
C 39	17.6	70.4	2269	6	US-10-750-185-48984	Sequence 48984, A
C 40	17.6	70.4	2339	6	US-10-750-185-51607	Sequence 51607, A
C 41	17.6	70.4	2338	6	US-10-750-185-29798	Sequence 29798, A
C 42	17.6	70.4	2476	6	US-10-986-501-86	Sequence 86, Appl1
C 43	17.6	70.4	2927	6	US-10-750-185-39842	Sequence 39842, A
C 44	17.6	70.4	2998	6	US-10-750-185-26597	Sequence 26597, A
C 45	17.6	70.4	3001	7	US-11-145-703-227	Sequence 227, App
C 46	17.6	70.4	3599	6	US-10-750-185-43445	Sequence 43445, A
C 47	17.6	70.4	28633	6	US-10-995-561-13341	Sequence 13341, A
C 48	17.6	70.4	46752	6	US-10-995-561-13410	Sequence 13410, A
C 49	17.6	70.4	55826	6	US-10-995-561-13256	Sequence 13256, A
C 50	17.6	70.4	63693	6	US-10-995-561-13269	Sequence 13269, A
C 51	17.6	70.4	65931	6	US-10-995-561-13254	Sequence 13254, A
C 52	17.6	70.4	68123	6	US-10-995-561-13348	Sequence 13348, A
C 53	17.6	70.4	85682	7	US-11-117-187-205	Sequence 205, App
C 54	17.6	70.4	119160	7	US-11-121-086-12	Sequence 12, Appl1
C 55	17.6	70.4	127340	7	US-11-112-908-35	Sequence 35, Appl1
C 56	17.6	70.4	142303	7	US-11-121-086-42	Sequence 42, Appl1
C 57	17.6	70.4	153142	7	US-11-121-086-27	Sequence 27, Appl1
C 58	17.6	70.4	157230	7	US-11-112-908-64	Sequence 64, Appl1
C 59	17.6	70.4	165883	7	US-11-112-908-18	Sequence 18, Appl1
C 60	17.6	70.4	169495	7	US-11-121-086-61	Sequence 61, Appl1
C 61	17.6	70.4	170508	7	US-11-112-908-62	Sequence 62, Appl1
C 62	17.6	70.4	172111	7	US-11-112-908-68	Sequence 68, Appl1
C 63	17.6	70.4	173115	7	US-11-112-908-65	Sequence 65, Appl1
C 64	17.6	70.4	184000	7	US-11-121-086-37	Sequence 37, Appl1
C 65	17.6	70.4	212766	7	US-11-121-086-95	Sequence 95, Appl1
C 66	17.6	70.4	217623	7	US-11-112-908-33	Sequence 33, Appl1
C 67	17.6	70.4	403287	6	US-10-995-561-13421	Sequence 13421, A
C 68	17.6	70.4	611587	7	US-11-117-187-209	Sequence 209, App
C 69	17.6	70.4	1082144	7	US-11-117-187-211	Sequence 211, App
C 70	17.4	69.6	25	7	US-11-121-849-24059	Sequence 24059, A
C 71	17.4	69.6	25	7	US-11-121-849-454515	Sequence 454515, A
C 72	17.4	69.6	201	6	US-10-995-561-24377	Sequence 24377, A
C 73	17.4	69.6	201	6	US-10-995-561-31677	Sequence 31677, A
C 74	17.4	69.6	201	6	US-10-995-561-56687	Sequence 56687, A
C 75	17.4	69.6	201	6	US-10-995-561-56697	Sequence 56697, A
C 76	17.4	69.6	201	6	US-10-995-561-56747	Sequence 56747, A
C 77	17.4	69.6	201	6	US-10-995-561-56773	Sequence 56773, A
C 78	17.4	69.6	201	6	US-10-995-561-57198	Sequence 57198, A
C 79	17.4	69.6	201	6	US-10-995-561-57199	Sequence 57199, A
C 80	17.4	69.6	201	6	US-10-995-561-57200	Sequence 57200, A
C 81	17.4	69.6	201	6	US-10-995-561-79839	Sequence 79839, A
C 82	17.4	69.6	201	6	US-10-995-561-79863	Sequence 79863, A
C 83	17.4	69.6	1176	6	US-10-750-185-29230	Sequence 29230, A
C 84	17.4	69.6	14670	6	US-10-995-561-13328	Sequence 13328, A
C 85	17.4	69.6	18097	6	US-10-995-561-13386	Sequence 13386, A
C 86	17.4	69.6	18895	6	US-10-995-561-13329	Sequence 13329, A
C 87	17.4	69.6	155826	6	US-10-995-561-13356	Sequence 13356, A
C 88	17.4	69.6	159138	6	US-10-995-561-13230	Sequence 13230, A
C 89	17.2	68.8	201	6	US-10-995-561-55548	Sequence 55548, A
C 90	17.2	68.8	201	6	US-10-995-561-55551	Sequence 55551, A
C 91	17.2	68.8	264	6	US-10-726-232A-22	Sequence 22, Appl1
C 92	17.2	68.8	600	6	US-10-750-185-3486	Sequence 3486, App
C 93	17.2	68.8	933	6	US-10-750-185-47253	Sequence 47253, A
C 94	17.2	68.8	1003	6	US-10-750-185-32132	Sequence 32132, A
C 95	17.2	68.8	1269	6	US-10-750-185-46263	Sequence 46263, A
C 96	17.2	68.8	1305	6	US-10-750-185-32792	Sequence 32792, A

```

c 97 17.2 68.8 1540 6 US-10-750-185-44033 Sequence 44033, A
c 98 17.2 68.8 1659 6 US-10-750-185-59617 Sequence 59617, A
c 99 17.2 68.8 1731 6 US-10-750-185-54846 Sequence 54846, A
c 100 17.2 68.8 1774 6 US-10-750-185-26214 Sequence 26214, A
c 101 17.2 68.8 1864 6 US-10-750-185-53720 Sequence 53720, A
c 102 17.2 68.8 1884 6 US-10-750-185-27841 Sequence 27841, A
c 103 17.2 68.8 2741 6 US-10-750-185-42880 Sequence 42880, A
c 104 17.2 68.8 3626 6 US-10-750-185-55135 Sequence 55135, A
c 105 17.2 68.8 3773 6 US-10-750-185-38377 Sequence 38377, A
c 106 17.2 68.8 4692 6 US-10-750-185-48511 Sequence 48511, A
c 107 17.2 68.8 124972 7 US-11-121-086-100 Sequence 100, App
c 108 17.2 68.8 171486 7 US-11-121-086-105 Sequence 105, App
c 109 17.2 68.8 179777 7 US-11-121-086-106 Sequence 106, App
c 110 17.2 68.8 185393 7 US-11-121-086-101 Sequence 101, App
c 111 17.2 68.8 197096 7 US-11-121-086-107 Sequence 107, App
c 112 17.2 68.8 285300 6 US-10-857-780-6 Sequence 6, App1
c 113 17.2 68.8 611587 7 US-11-117-187-209 Sequence 209, App
c 114 17.2 68.8 112500 6 US-10-995-561-13286 Sequence 13286, A
c 115 17.2 68.8 201 6 US-10-995-561-21428 Sequence 21428, A
c 116 17.2 68.8 201 6 US-10-995-561-28005 Sequence 28005, A
c 117 17.2 68.8 201 6 US-10-995-561-28363 Sequence 28363, A
c 118 17.2 68.8 201 6 US-10-995-561-36062 Sequence 36062, A
c 119 17.2 68.8 201 6 US-10-995-561-42372 Sequence 42372, A
c 120 17.2 68.8 201 6 US-10-995-561-44196 Sequence 44196, A
c 121 17.2 68.8 201 6 US-10-995-561-61010 Sequence 61010, A
c 122 17.2 68.8 201 6 US-10-995-561-61171 Sequence 61171, A
c 123 17.2 68.8 201 6 US-10-995-561-68245 Sequence 68245, A
c 124 17.2 68.8 201 6 US-10-995-561-75746 Sequence 75746, A
c 125 17.2 68.8 201 6 US-10-995-561-84522 Sequence 84522, A
c 126 17.2 68.8 861 6 US-10-750-185-47370 Sequence 47370, A
c 127 17.2 68.8 916 6 US-10-750-185-41319 Sequence 41319, A
c 128 17.2 68.8 1020 6 US-10-750-185-54011 Sequence 54011, A
c 129 17.2 68.8 1076 6 US-10-750-185-29549 Sequence 29549, A
c 130 17.2 68.8 1374 6 US-10-750-185-47635 Sequence 47635, A
c 131 17.2 68.8 1425 6 US-10-750-185-63170 Sequence 63170, A
c 132 17.2 68.8 1459 6 US-10-750-185-49889 Sequence 49889, A
c 133 17.2 68.8 1536 6 US-10-750-185-60649 Sequence 60649, A
c 134 17.2 68.8 1538 6 US-10-750-185-64129 Sequence 64129, A
c 135 17.2 68.8 1570 6 US-10-750-185-32404 Sequence 32404, A
c 136 17.2 68.8 1748 6 US-10-750-185-24820 Sequence 24820, A
c 137 17.2 68.8 1968 6 US-10-750-185-25344 Sequence 25344, A
c 138 17.2 68.8 3194 6 US-10-750-185-32017 Sequence 32017, A
c 139 17.2 68.8 3330 6 US-10-750-185-61867 Sequence 61867, A
c 140 17.2 68.8 9808 7 US-11-021-441-33 Sequence 33, App1
c 141 17.2 68.0 16175 6 US-10-995-561-13399 Sequence 13399, A
c 142 17.2 68.0 16963 6 US-10-995-561-13467 Sequence 13467, A
c 143 17.2 68.0 20773 6 US-10-995-561-13222 Sequence 13222, A
c 144 17.2 68.0 33014 7 US-11-077-716-1 Sequence 1, App1
c 145 17.2 68.0 34555 6 US-10-623-155-479 Sequence 479, App
c 146 17.2 68.0 35656 6 US-10-860-436-1 Sequence 1, App1
c 147 17.2 68.0 35893 6 US-10-860-436-2 Sequence 2, App1
c 148 17.2 68.0 40000 6 US-10-995-561-13513 Sequence 13513, A
c 149 17.2 68.0 88116 6 US-10-995-561-13351 Sequence 13351, A
c 150 17.2 68.0 95832 6 US-10-995-561-13273 Sequence 13273, A
```

ALIGNMENTS

```

RESULT 1
US-10-750-185-20178/c
; Sequence 20178, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
```

```

; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 20178
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Bovine MMBT18831
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: n is any nucleotide
US-10-750-185-20178
```

```

Query Match 79.2%; Score 19.8; DB 6; Length 596;
Best Local Similarity 91.3%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 1 AAAAAAAAAAGCATGATTGTGAC 23
Db 80 AAAAAAAAAAGCAAGTTGTGAC 58
```

```

RESULT 2
US-10-750-185-39819/c
; Sequence 39819, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 39819
; LENGTH: 875
; TYPE: DNA
; ORGANISM: Bovine 19866880826868
US-10-750-185-39819

Query Match 79.2%; Score 19.8; DB 6; Length 875;
Best Local Similarity 91.3%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 1 AAAAAAAAAAGCATGATTGTGAC 23
Db 359 AAAAAAAAAAGCAAGTTGTGAC 337

RESULT 3
US-11-121-086-8
; Sequence 8, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
```


NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 8
LENGTH: 246960
TYPE: DNA
ORGANISM: Homo sapiens
US-11-121-086-8

Query Match 77.6%; Score 19.4; DB 7; Length 246960;
Best Local Similarity 95.2%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTC 21
Db 101580 AAAAAAAAAAGCATGATTGTC 101600

RESULT 4
US-10-750-185-35652
Sequence 35652, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
PRIOR FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35652
LENGTH: 737
TYPE: DNA
ORGANISM: Bovine 19866881468215
US-10-750-185-35652

Query Match 76.8%; Score 19.2; DB 6; Length 737;
Best Local Similarity 87.5%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTC 24
Db 351 AAAAAAAAAAGCAAAATTGTGACA 374

RESULT 5
US-10-750-185-61429/C
Sequence 61429, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
PRIOR FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 61429

LENGTH: 2275
TYPE: DNA
ORGANISM: Bovine 19866880595799
US-10-750-185-61429

Query Match 76.8%; Score 19.2; DB 6; Length 2275;
Best Local Similarity 87.5%; Pred. No. 44;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTC 24
Db 362 AAAAAAAAAACATGTTTCGACA 339

RESULT 6
US-10-750-185-53588
Sequence 53588, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
PRIOR FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 53588
LENGTH: 5571
TYPE: DNA
ORGANISM: Bovine 19866880753767
US-10-750-185-53588

Query Match 75.2%; Score 18.8; DB 6; Length 5571;
Best Local Similarity 90.9%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTC 22
Db 3736 AAAAAAAAAAGCATCATATGA 3757

RESULT 7
US-10-750-185-40699
Sequence 40699, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
PRIOR FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 40699
LENGTH: 2049
TYPE: DNA
ORGANISM: Bovine 19866881376320

US-10-750-185-40699

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 2049;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
DB 418 AAAAAAAAAAGTGTGATCTGCACCC 442

RESULT 8

US-10-750-185-56281
; Sequence 56281, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56281
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Bovine 1986680800952
US-10-750-185-56281

Query Match
Best Local Similarity 84.0%; Score 18.6; DB 6; Length 2477;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
DB 2302 AAAAAAAAAATCTCATTTTGACAC 2326

RESULT 9

US-10-750-185-45175/C
; Sequence 45175, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45175
; LENGTH: 3138
; TYPE: DNA
; ORGANISM: Bovine 19866808054433
US-10-750-185-45175

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 3138;

Best Local Similarity 84.0%; Pred. No. 80;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
DB 651 AATMAAAGAAATTAATCTGCAC 627

RESULT 10

US-11-121-086-30/C
; Sequence 30, Application US/11121086
; Publication No. US2005026459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138 6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 30
; LENGTH: 158692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-30

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 7; Length 158692;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
DB 138902 AAAAAAAAAAGATTAATGATACAC 138878

RESULT 11

US-11-112-908-60
; Sequence 60, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: HARRIS, Cole
; APPLICANT: DAVIS, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 60
; LENGTH: 171427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-60

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 7; Length 171427;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
DB 165790 AAAAAAAAAATCATGAAGTGATAC 165814

RESULT 12
US-10-995-561-13244
; Sequence 13244, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13244
; LENGTH: 222094
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13244

Query Match 74.4%; Score 18.6; DB 6; Length 222094;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
Db 111412 AAAAAAAAAAGCATGATTGTGACAC 111436

RESULT 13
US-11-121-086-61/C
; Sequence 61, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138,6000-00000
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
; LENGTH: 169495
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (70072)..(70171)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (139457)..(157244)
; OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-61

Query Match 73.6%; Score 18.4; DB 7; Length 169495;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGT 20
Db 82652 AAAAAAAAAAGCATGATTGT 82633

RESULT 14
US-10-750-185-47195
; Sequence 47195, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47195
; LENGTH: 1282
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-47195

Query Match 72.8%; Score 18.2; DB 6; Length 1282;
Best Local Similarity 87.0%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGAC 23
Db 1211 AAAAAAAAAAGCATGATTGTGCC 1233

RESULT 15
US-11-112-908-29
; Sequence 29, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 29
; LENGTH: 131855
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-29

Query Match 72.8%; Score 18.2; DB 7; Length 131855;
Best Local Similarity 87.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGCATGATTGTGACAC 25
Db 23371 AAAAAAAAAAGCATGATTGTGACAC 23393

RESULT 16
US-11-112-908-26
; Sequence 26, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.

```
;; TITLE OF INVENTION: Breast Cancer Biomarkers
;; FILE REFERENCE: 04-164-US
;; CURRENT APPLICATION NUMBER: US/11/112,908
;; PRIOR FILING DATE: 2005-04-22
;; PRIOR APPLICATION NUMBER: US 60/564,758
;; PRIOR FILING DATE: 2004-04-23
;; PRIOR APPLICATION NUMBER: US 60/575,978
;; PRIOR FILING DATE: 2004-06-01
;; PRIOR APPLICATION NUMBER: US 60/631,702
;; PRIOR FILING DATE: 2004-11-30
;; PRIOR APPLICATION NUMBER: US 60/633,826
;; PRIOR FILING DATE: 2004-12-07
;; NUMBER OF SEQ ID NOS: 511
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 26
;; LENGTH: 150173
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-11-112-908-26
```

```
Query Match          72.8%; Score 18.2; DB 7; Length 150173;
Best Local Similarity 87.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 AAAAAAAAAAGCATGATTGTGACAC 25
Db      93075 AAAAAAAAAATATGATTGGACAC 93097
```

```
RESULT 17
US-11-112-908-28
;; Sequence 28, Application US/11/112908
;; Publication No. US20050260659A1
;; GENERAL INFORMATION:
;; APPLICANT: Harris, Cole
;; APPLICANT: Harris, Cole
;; TITLE OF INVENTION: Breast Cancer Biomarkers
;; FILE REFERENCE: 04-164-US
;; CURRENT APPLICATION NUMBER: US/11/112,908
;; CURRENT FILING DATE: 2005-04-22
;; PRIOR APPLICATION NUMBER: US 60/564,758
;; PRIOR FILING DATE: 2004-04-23
;; PRIOR APPLICATION NUMBER: US 60/575,978
;; PRIOR FILING DATE: 2004-06-01
;; PRIOR APPLICATION NUMBER: US 60/631,702
;; PRIOR FILING DATE: 2004-11-30
;; PRIOR APPLICATION NUMBER: US 60/633,826
;; PRIOR FILING DATE: 2004-12-07
;; NUMBER OF SEQ ID NOS: 511
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 28
;; LENGTH: 166020
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-11-112-908-28
```

```
Query Match          72.8%; Score 18.2; DB 7; Length 166020;
Best Local Similarity 87.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 AAAAAAAAAAGCATGATTGTGACAC 25
Db      37625 AAAAAAAAAATATGATTGGACAC 37647
```

```
RESULT 18
US-11-112-908-27
;; Sequence 27, Application US/11/112908
;; Publication No. US20050260659A1
;; GENERAL INFORMATION:
;; APPLICANT: Harris, Cole
;; APPLICANT: Harris, Cole
;; TITLE OF INVENTION: Breast Cancer Biomarkers
```

```
;; FILE REFERENCE: 04-164-US
;; CURRENT APPLICATION NUMBER: US/11/112,908
;; PRIOR FILING DATE: 2005-04-22
;; PRIOR APPLICATION NUMBER: US 60/564,758
;; PRIOR FILING DATE: 2004-04-23
;; PRIOR APPLICATION NUMBER: US 60/575,978
;; PRIOR FILING DATE: 2004-06-01
;; PRIOR APPLICATION NUMBER: US 60/631,702
;; PRIOR FILING DATE: 2004-11-30
;; PRIOR APPLICATION NUMBER: US 60/633,826
;; PRIOR FILING DATE: 2004-12-07
;; NUMBER OF SEQ ID NOS: 511
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 27
;; LENGTH: 171247
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-11-112-908-27
```

```
Query Match          72.8%; Score 18.2; DB 7; Length 171247;
Best Local Similarity 87.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 AAAAAAAAAAGCATGATTGTGACAC 25
Db      87584 AAAAAAAAAATATGATTGGACAC 87606
```

```
RESULT 19
US-11-121-086-31
;; Sequence 31, Application US/11/121086
;; Publication No. US20050266459A1
;; GENERAL INFORMATION:
;; APPLICANT: POULSEN, TIM S.
;; APPLICANT: NIELSEN, KIRSTEN V.
;; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
;; FILE REFERENCE: 09138.6000-00000
;; CURRENT APPLICATION NUMBER: US/11/121,086
;; CURRENT FILING DATE: 2005-05-04
;; PRIOR APPLICATION NUMBER: 60/567,570
;; PRIOR FILING DATE: 2004-05-04
;; NUMBER OF SEQ ID NOS: 107
;; SOFTWARE: PatentIn version 3.3
;; SEQ ID NO 31
;; LENGTH: 218821
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: modified_base
;; LOCATION: (106949)..(106949)
;; OTHER INFORMATION: a, c, g, t, unknown or other
;; FEATURE:
;; NAME/KEY: modified_base
;; LOCATION: (110322)..(110324)
;; OTHER INFORMATION: a, c, g, t, unknown or other
;; FEATURE:
;; NAME/KEY: modified_base
;; LOCATION: (115133)..(115133)
;; OTHER INFORMATION: a, c, g, t, unknown or other
;; FEATURE:
;; NAME/KEY: modified_base
;; LOCATION: (139059)..(139158)
;; OTHER INFORMATION: a, c, g, t, unknown or other
;; NAME/KEY: modified_base
;; LOCATION: (131300)..(131300)
;; OTHER INFORMATION: a, c, g, t, unknown or other
;; NAME/KEY: modified_base
;; LOCATION: (157740)..(157740)
;; OTHER INFORMATION: a, c, g, t, unknown or other
;; NAME/KEY: modified_base
```

LOCATION: (157777)..(157777)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157900)..(157900)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157919)..(157919)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157926)..(157926)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158094)..(158094)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158138)..(158138)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158193)..(158193)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158241)..(158242)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158259)..(158259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158278)..(158278)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158295)..(158295)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158740)..(158839)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158929)..(158929)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163528)..(163530)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163550)..(163550)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163765)..(163765)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164000)..(164000)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164047)..(164047)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164084)..(164084)

OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167233)..(167233)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167236)..(167236)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167238)..(167238)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170247)..(170247)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170249)..(170250)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170252)..(170253)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170259)..(170259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170262)..(170263)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170266)..(170266)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174470)..(174470)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174472)..(174472)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174474)..(174474)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179059)..(179060)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179064)..(179064)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197001)..(197001)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197004)..(197005)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197007)..(197007)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (200349)..(200349)
OTHER INFORMATION: a, c, g, t, unknown or other

```
FEATURE:
NAME/KEY: modified_base
LOCATION: (200351)..(200351)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (212425)..(212426)
OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-31
```

```
Query Match          72.8%; Score 18.2; DB 7; Length 218821;
Best Local Similarity 87.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTGAC 23
          |||
Db      32967 AAAAAAAAAAGCATGACGGGCG 32989
```

```
RESULT 20
US-10-995-561-57202
; Sequence 57202, Application US/10995561
; Publication No. US20050272054A1
```

```
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57202
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-57202
```

```
Query Match          72.0%; Score 18; DB 6; Length 201;
Best Local Similarity 90.0%; Pred. No. 92;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGT 20
          |||
Db      101 AAAAAAAAAAGCATATTGT 120
```

```
RESULT 21
US-10-995-561-13331
; Sequence 13331, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13331
; LENGTH: 98716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13331
```

```
Query Match          72.0%; Score 18; DB 6; Length 98716;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGT 20
```

```
Db      63658 AAAAAAAAAAGCATATTGT 63677
```

```
RESULT 22
US-10-995-561-83376/c
; Sequence 83376, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83376
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-83376
```

```
Query Match          71.2%; Score 17.8; DB 6; Length 201;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTG 21
          |||
Db      75 AAAAAAAAAAGCATGATGCTG 55
```

```
RESULT 23
US-10-750-185-45080
; Sequence 45080, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 45080
; LENGTH: 3734
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-45080
```

```
Query Match          71.2%; Score 17.8; DB 6; Length 3734;
Best Local Similarity 90.5%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAAAGCATGATTGTG 21
          |||
Db      2160 AAAAAAAAAAGCATGATTCTG 2180
```

```
RESULT 24
US-10-995-561-13506/c
; Sequence 13506, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
```

```

; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13506
; LENGTH: 17797
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(17797)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13506

Query Match 71.2%; Score 17.8; DB 6; Length 17797;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGTTGTG 21
Db 12362 AAAAAAAAAAGTATGATGTTG 12342

RESULT 25
US-11-155-492-1/C
; Sequence 1, Application US/11155492
; Publication No. US20050266479A1
; GENERAL INFORMATION:
; APPLICANT: Hazan, Jamil
; TITLE OF INVENTION: CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4
; TITLE OF INVENTION: GENE RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL
; TITLE OF INVENTION: DOMINANT SPASTIC PARAPLEGIA
; FILE REFERENCE: R-341894
; CURRENT APPLICATION NUMBER: US/11/155,492
; CURRENT FILING DATE: 2005-06-20
; PRIOR APPLICATION NUMBER: US/09/830,902
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: FR 99 11097
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: PCT/FR00/02433
; PRIOR FILING DATE: 2000-09-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 1
; LENGTH: 110000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (9932)...(10471)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (10472)...(33718)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (33719)...(33805)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (33806)...(35748)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (35749)...(35832)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (35833)...(45022)
; FEATURE:
; NAME/KEY: exon
```

```

; LOCATION: (45023)...(45118)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (45119)...(60863)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (60864)...(61051)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (61052)...(61927)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (61928)...(62061)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (62062)...(62344)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (62345)...(62438)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (62439)...(73173)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (73174)...(73248)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (73249)...(74633)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (74634)...(74705)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (74706)...(82788)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (82789)...(82864)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (82865)...(83102)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (83103)...(83194)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (83195)...(83334)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (83335)...(83414)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (83415)...(88129)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (88130)...(88172)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (88173)...(89561)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (89562)...(89641)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (89642)...(91162)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (91163)...(91233)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (91234)...(93443)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (93444)...(93484)
```

```

; FEATURE:
; NAME/KEY: introm
; LOCATION: (93485) ... (100599)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (100600) ... (102009)
US-11-155-492-1

```

Query Match	71.2%	Score 17.8;	DB 7;	Length 110000;
Best Local Similarity	90.5%	Pred. No. 2.8e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

```

Qy      1 AAAAAAAAAAGCATGATTGTG 21
          |||||
Db      109689 AAAAAAAAAAGCAGATTATG 109665

```

```

RESULT 26
US-11-121-086-58/c
; Sequence 58, Application US/11121086
; Publication No. US2005026649A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 58
; LENGTH: 180654
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-58

```

Query Match Similarity	71.2%	Score 17.8	DB 7	Length 180654
Best Local Similarity	90.5%	Pred. No. 2.9e+02		
Matches 19	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	1 AAAAAAAAAAGCATGATTGTG	21		
DB	124208 AAAAAAAAAAGCATGATTGTG	124188		

```

RESULT 27
US-10-995-561-31582
; Sequence 31582, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31582
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-31582

```

Query Match	70.4%	Score 17.6;	DB 6;	Length 201;
Best Local Similarity	83.3%;	Pred. No. 1.3e+02;		
Matches	20;	Conservative	4;	Indels 0; Gaps 0
QY	1	AAAAAAAAAGCATGATTGTGACA	24	

Db 92 AAAAAAAAAATCATGCTAGTCCA 115

RESULT 28
US-10-995-561-31727
; Sequence 31727, Application US/10995561
; Publication No. US20050272054A1

```

/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 31727
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-31727

```

	Query Match	70.4%	Score 17.6	ID 6	Length 201
	Best Local Similarity	83.3%	Pred. No. 1.3e+02		
	Matches	20	Conservative	0	Mismatches 4; Indels 0; Gaps 0
QY	1	AAAAAAAAAAGCATGATTGTGCACA	24		
DB	158	AAAAAAAAAAATCATCTAGTGCCCA	181		

```

RESULT 29
US-10-995-561-31728
/ Sequence 31728, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 31728
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-31728

```

	Query Match	70.4%	Score 17.6;	DB 6;	Length 201;
	Best Local Similarity	83.3%	Pred. No. 1.3e+02;		
	Matches	20;	Conservative	0;	Mismatches 4; Indels 0; Gaps 0
Qy	1	AAAAAAAAAGCATGATTGTGCACA	24		
Dd	177	AAAAAAAAAATCATCTAGTGCCA	200		

```

RESULT 30
US-10-995-561-58952/C
; Sequence 58952, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24

```


NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 58952
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-58952

Query Match 70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 177 AAAAAAAAAAGCATGATTGTGACAAA 154

RESULT 31
US-10-995-561-69364
Sequence 69364, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: C0001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 69364
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-69364

Query Match 70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 93 AAAAAAAAAAGCATGATTGTGACACA 116

RESULT 32
US-10-750-185-618
Sequence 618, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 618
LENGTH: 600
TYPE: DNA
ORGANISM: Bovine
US-10-750-185-618

Query Match 70.4%; Score 17.6; DB 6; Length 600;

Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 353 AAAAAAAAAAGCATCTCTGTGACA 376

RESULT 33
US-10-750-185-48111
Sequence 48111, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48111
LENGTH: 979
TYPE: DNA
ORGANISM: Bovine
US-10-750-185-48111

Query Match 70.4%; Score 17.6; DB 6; Length 979;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 886 AAAAAAAAAAGTATGTTTGAGACA 909

RESULT 34
US-10-750-185-45903
Sequence 45903, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 45903
LENGTH: 1049
TYPE: DNA
ORGANISM: Bovine
US-10-750-185-45903

Query Match 70.4%; Score 17.6; DB 6; Length 1049;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
DB 837 AAAAAAAAAAGCATGATTGTGACA 860

RESULT 35

US-10-750-185-59829/C
Sequence 59829, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 59829
LENGTH: 1193
TYPE: DNA
ORGANISM: Bovine 19866881354501
US-10-750-185-59829

Query Match 70.4%; Score 17.6; DB 6; Length 1193;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
DB 396 AAAAAAAAAAGCATGATTGTGACA 373

RESULT 36

US-10-750-185-30513/C
Sequence 30513, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 30513
LENGTH: 1328
TYPE: DNA
ORGANISM: Bovine 19866881384544
US-10-750-185-30513

Query Match 70.4%; Score 17.6; DB 6; Length 1328;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
DB 422 AAAAAAAAAAGCATGATTGTGACA 399

RESULT 37

US-10-750-185-64253
Sequence 64253, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 64253
LENGTH: 1760
TYPE: DNA
ORGANISM: Bovine 19866880644297
US-10-750-185-64253

Query Match 70.4%; Score 17.6; DB 6; Length 1760;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
DB 845 AAAAAAAAAAGCATGATTGTGACA 868

RESULT 38

US-10-750-185-64736
Sequence 64736, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 64736
LENGTH: 2051
TYPE: DNA
ORGANISM: Bovine 19866880837218
US-10-750-185-64736

Query Match 70.4%; Score 17.6; DB 6; Length 2051;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
DB 332 AAAAAAAAAAGCATGATTGTGACA 355

RESULT 39

```
US-10-750-185-48984
; Sequence 48984, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48984
; LENGTH: 2269
; TYPE: DNA
; ORGANISM: Bovine 1986681340823
US-10-750-185-48984

Query Match          70.4%; Score 17.6; DB 6; Length 2269;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATGTGACA 24
Db 1050 AAAAAAAAAAGCATGATTTTAAAA 1073

RESULT 40
US-10-750-185-51607/c
; Sequence 51607, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51607
; LENGTH: 2329
; TYPE: DNA
; ORGANISM: Bovine 1986680449489
US-10-750-185-51607

Query Match          70.4%; Score 17.6; DB 6; Length 2329;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATGTGACA 24
Db 549 AAAAAAAAAAGCATATTGTACACA 526

RESULT 41
US-10-750-185-29798
; Sequence 29798, Application US/10750185
; Publication No. US20050260603A1
```

```
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29798
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Bovine 1986680352450
US-10-750-185-29798

Query Match          70.4%; Score 17.6; DB 6; Length 2334;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATGTGACA 24
Db 1784 AAAAAAAAAAGCATTTATGACA 1807

RESULT 42
US-10-986-501-86/c
; Sequence 86, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; PRIOR FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 86
; LENGTH: 2476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (853)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
```

```
/ NAME/KEY: misc.feature
/ LOCATION: (2227)
/ OTHER INFORMATION: n equals a,t,g, or c
US-10-986-501-86

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 6; Length 2476;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGATTGTGACAC 25
DB 1718 AATAAGAGCATGTTTGAGACAC 1695

RESULT 43
US-10-750-185-39842/c
/ Sequence 39842, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ PRIOR FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: Patent version 3.1
/ SEQ ID NO 39842
/ LENGTH: 2927
/ TYPE: DNA
/ ORGANISM: Bovine 19866880775971
US-10-750-185-39842

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 6; Length 2927;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 148 AAAAAATAAGATGATATATGACA 125

RESULT 44
US-10-750-185-26597
/ Sequence 26597, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ PRIOR FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: Patent version 3.1
/ SEQ ID NO 26597
/ LENGTH: 2998
/ TYPE: DNA
/ ORGANISM: Bovine 19866881169843
```

```
US-10-750-185-26597

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 6; Length 2998;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 1904 AAAAAAGCAATCTGTATACA 1927

RESULT 45
US-11-145-703-227/c
/ Sequence 227, Application US/11145703
/ Publication No. US20050260667A1
/ GENERAL INFORMATION:
/ APPLICANT: Cohen, Daniel
/ APPLICANT: Blumenfeld, Marta
/ APPLICANT: Chumakov, Ilya
/ APPLICANT: Bouguetelat, Lydie
/ APPLICANT: Bihain, Bernard
/ APPLICANT: Essioux, Laurent
/ TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
/ FILE REFERENCE: 53.US16.DIV
/ CURRENT APPLICATION NUMBER: US/11/145,703
/ PRIOR FILING DATE: 2000-03-30
/ PRIOR APPLICATION NUMBER: US/10/147,603
/ PRIOR FILING DATE: 2000-03-30
/ PRIOR APPLICATION NUMBER: 09/539,333
/ PRIOR FILING DATE: 2000-03-30
/ PRIOR APPLICATION NUMBER: US 60/126,903
/ PRIOR FILING DATE: 1999-03-30
/ PRIOR APPLICATION NUMBER: US 60/131,971
/ PRIOR FILING DATE: 1999-04-30
/ PRIOR APPLICATION NUMBER: US 60/132,065
/ PRIOR FILING DATE: 1999-04-30
/ PRIOR APPLICATION NUMBER: US 60/143,928
/ PRIOR FILING DATE: 1999-07-14
/ PRIOR APPLICATION NUMBER: US 60/145,915
/ PRIOR FILING DATE: 1999-07-27
/ PRIOR APPLICATION NUMBER: US 60/146,453
/ PRIOR FILING DATE: 1999-07-29
/ PRIOR APPLICATION NUMBER: US 60/146,452
/ PRIOR FILING DATE: 1999-07-29
/ PRIOR APPLICATION NUMBER: US 60/162,288
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 234
/ SOFTWARE: Patent.pm
/ SEQ ID NO 227
/ LENGTH: 3001
/ TYPE: DNA
/ ORGANISM: Homo Sapiens
/ FEATURE:
/ NAME/KEY: allele
/ LOCATION: 1501
/ OTHER INFORMATION: 99-27365-421 : polymorphic base C or T
/ FEATURE:
/ NAME/KEY: misc binding
/ LOCATION: 1482_1500
/ OTHER INFORMATION: 99-27365-421.mis1
/ FEATURE:
/ NAME/KEY: misc binding
/ LOCATION: 1502_1521
/ OTHER INFORMATION: 99-27365-421.mis2, complement
/ FEATURE:
/ NAME/KEY: primer bind
/ LOCATION: 1081..1099
/ OTHER INFORMATION: upstream amplification primer
/ FEATURE:
/ NAME/KEY: primer bind
/ LOCATION: 1590..1609
/ OTHER INFORMATION: downstream amplification primer, complement
```

```
FEATURE:
; NAME/KEY: misc binding
; LOCATION: 1489_1513
; OTHER INFORMATION: 99-27365-421 probe
US-11-145-703-227
```

```
Query Match      70.4%; Score 17.6; DB 7; Length 3001;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 452 AAAAAAAAAAGCATGATTGTGAAA 429
```

```
RESULT 46
US-10-750-185-43445
; Sequence 43445, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43445
; LENGTH: 3599
; TYPE: DNA
; ORGANISM: Bovine 19866880568392
US-10-750-185-43445
```

```
Query Match      70.4%; Score 17.6; DB 6; Length 3599;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 2118 AAAAAAAAAAGATTGTGTGACA 2141
```

```
RESULT 47
US-10-995-561-13341/C
; Sequence 13341, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13341
; LENGTH: 28693
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13341
```

```
Query Match      70.4%; Score 17.6; DB 6; Length 28693;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 22787 AAAAAAAAAAGATGAGTGAGAAA 22764
```

```
RESULT 48
US-10-995-561-13410
; Sequence 13410, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13410
; LENGTH: 46752
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13410
```

```
Query Match      70.4%; Score 17.6; DB 6; Length 46752;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 45723 AAAAAAAAAAGCTTGATTGACACA 45746
```

```
RESULT 49
US-10-995-561-13256
; Sequence 13256, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13256
; LENGTH: 55826
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13256
```

```
Query Match      70.4%; Score 17.6; DB 6; Length 55826;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 51873 AAAAAAAAAATCATGCTAGTGCCA 51896
```

```
RESULT 50
US-10-995-561-13269
; Sequence 13269, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
```

```
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13269
; LENGTH: 63693
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13269
```

```
Query Match          70.4%; Score 17.6; DB 6; Length 63693;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAAGCATGATTTGTGACA 24
          |||||
Db      7054 AAAAAAAAAAGCTTTATTGAGATA 7077
```

Search completed: December 14, 2005, 11:40:37
Job time : 186.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

qM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-1

Perfect score: 25
Sequence: 1 aaaaaaaaaatcgaacaatact 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl:*
1: gb_ha:*
2: gb_in:*
3: gb_env:*
4: gb_ov:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_ses:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	40899	8	CH19F14121
C 2	25	100.0	149964	8	AC104686
C 3	23.4	93.6	49613	8	AL359186
C 4	23.4	93.6	59765	8	AC005179
C 5	23.4	93.6	90007	8	AL139234
C 6	23.4	93.6	108736	8	CNS01RGR
C 7	23.4	93.6	111002	8	AL139179
C 8	23.4	93.6	112862	15	AL139143
C 9	23.4	93.6	112862	14	AC025781
C 10	23.4	93.6	114835	8	AP001628
C 11	23.4	93.6	129402	8	AC011510
C 12	23.4	93.6	135754	8	AC125612
C 13	23.4	93.6	142728	8	HS0792G4
C 14	23.4	93.6	143334	8	AL137881
C 15	23.4	93.6	151875	8	AL136128
C 16	23.4	93.6	154957	8	AC019270
C 17	23.4	93.6	161659	14	AC147311
C 18	23.4	93.6	161841	8	AC135279

C 19	23.4	93.6	163554	14	AC021786
C 20	23.4	93.6	163731	8	AC008949
C 21	23.4	93.6	166517	14	AL591663
C 22	23.4	93.6	168833	14	AC021706
C 23	23.4	93.6	169993	8	AL353616
C 24	23.4	93.6	171732	8	AC022844
C 25	23.4	93.6	172779	8	AL451107
C 26	23.4	93.6	173491	14	AC073894
C 27	23.4	93.6	189791	14	AC152354
C 28	23.4	93.6	196831	14	AC068681
C 29	23.4	93.6	224450	14	AC016311
C 30	23.4	93.6	224450	14	AC016311
C 31	23.4	93.6	242704	8	BS000230
C 32	23.4	93.6	340000	8	AP001747
C 33	22.4	89.6	14026	8	AC092209
C 34	22.4	89.6	68419	14	AC016014
C 35	22.4	89.6	84791	8	AC113340
C 36	22.4	89.6	134599	8	AC011385
C 37	22.4	89.6	147263	14	AC142718
C 38	22.4	89.6	152010	14	CR936486
C 39	22.4	89.6	162753	14	AC144197
C 40	22.4	89.6	163408	14	AC091389
C 41	22.4	89.6	169212	8	AP002392
C 42	22.4	89.6	178088	14	AC073645
C 43	22.4	89.6	237547	14	AC144198
C 44	22.4	88.0	7300	8	AB000462
C 45	22.4	88.0	22970	8	HS1247F6
C 46	22.4	88.0	149969	8	HSJ323A2
C 47	22.4	88.0	160111	14	AP000780
C 48	22.4	88.0	196476	14	AC055890
C 49	22.4	88.0	199255	14	AC084195
C 50	21.8	87.2	460	10	BV304791
C 51	21.8	87.2	609	10	BV373272
C 52	21.8	87.2	615	10	BV621513
C 53	21.8	87.2	760	10	BV629695
C 54	21.8	87.2	862	10	BV467903
C 55	21.8	87.2	6162	6	AX356490
C 56	21.8	87.2	7145	6	AX279989
C 57	21.8	87.2	7145	6	AX356409
C 58	21.8	87.2	7928	6	AX251751
C 59	21.8	87.2	7928	6	AX344165
C 60	21.8	87.2	7928	6	AX344997
C 61	21.8	87.2	7928	6	AX348556
C 62	21.8	87.2	9311	8	AC148675
C 63	21.8	87.2	29718	8	HS310H5
C 64	21.8	87.2	34956	8	AY398667
C 65	21.8	87.2	37805	2	AF125964
C 66	21.8	87.2	39188	8	AC004754
C 67	21.8	87.2	43594	8	AC093144
C 68	21.8	87.2	44226	8	AY130859
C 69	21.8	87.2	50409	14	AC108001
C 70	21.8	87.2	51825	8	AC118558
C 71	21.8	87.2	57453	14	AC110060
C 72	21.8	87.2	64562	14	AL135900_3
C 73	21.8	87.2	75120	8	AC004666
C 74	21.8	87.2	76474	14	AC139013
C 75	21.8	87.2	77223	14	AC033228
C 76	21.8	87.2	78011	8	HS48F12
C 77	21.8	87.2	80327	14	AC135619
C 78	21.8	87.2	80993	8	AC128656
C 79	21.8	87.2	81590	14	AC138338
C 80	21.8	87.2	82020	8	AC090448
C 81	21.8	87.2	83554	14	AC135997
C 82	21.8	87.2	89603	8	HSU63334
C 83	21.8	87.2	92893	8	AC020610
C 84	21.8	87.2	96638	8	AC092545
C 85	21.8	87.2	97037	8	AC004973
C 86	21.8	87.2	101954	8	AL161637
C 87	21.8	87.2	102562	14	AL109516
C 88	21.8	87.2	102743	8	AL645507
C 89	21.8	87.2	104269	8	AC105916
C 90	21.8	87.2	105960	8	HS0209A6
C 91	21.8	87.2	106514	8	AC079463

AC021786 Homo sapi
AC008949 Homo sapi
AL591663 Homo sapi
AC021706 Homo sapi
AL353616 Homo sapi
AC022844 Homo sapi
AL451107 Homo sapi
AC073894 Homo sapi
AC152354 Pan trogl
AC068681 Homo sapi
AC016311 Homo sapi
AC016311 Homo sapi
BS000230 Pan trogl
AP001747 Homo sapi
AC092209 Homo sapi
AC016014 Homo sapi
AC113340 Homo sapi
AC011385 Homo sapi
AC142718 Macaca mu
CR936486 Danio rer
AC144197 Macaca mu
AC091389 Homo sapi
AP002392 Homo sapi
AC073645 Homo sapi
AC144198 Macaca mu
AB000462 Homo sapi
Z68279 Human DNA s
AL121750 Human DNA
AP000780 Homo sapi
AC055890 Homo sapi
AC084195 Homo sapi
BV304791 S236P6314
BV373272 S231P6598
BV621513 S215P6603
BV629695 S216P6602
BV467903 G591P6000
AX356490 Sequence
AX279989 Sequence
AX356409 Sequence
AX251751 Sequence
AX344165 Sequence
AX344997 Sequence
AX348556 Sequence
AC148675 Macaca mu
Z69705 Human DNA s
AF125964 Caenorhab
AC004754 Homo sapi
AC093144 Homo sapi
AY130859 Homo sapi
AC108001 Homo sapi
AC118558 Homo sapi
AC110060 Homo sapi
Continuation (4 of
AC004666 Homo sapi
AC139013 Homo sapi
AC033228 Homo sapi
AL008012 Human DNA
AC135619 Homo sapi
AC128656 Homo sapi
AC138338 Homo sapi
AC090448 Homo sapi
AC135997 Homo sapi
U63334 Human KIT p
AC020610 Homo sapi
AC092545 Homo sapi
AC004973 Homo sapi
AL161637 Human DNA
AL109516 Homo sapi
AL645507 Homo sapi
AC105916 Homo sapi
AL035916 Homo sapi
AC079463 Homo sapi

misc_feature 2109..2261
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
...; putative"
exon 2109..2260
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=4
exon 2434..2596
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=5
misc_feature 2434..2596
/note="predicted exon, graill2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"
2435..2599
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
...; putative"
misc_feature 3675..3775
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
...; putative"
exon 3678..3775
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=6
misc_feature 3678..3775
/note="predicted exon, graill2exons human_1.3; frame=2,
forward strand, quality=excellent; putative"
3867..3939
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=7
misc_feature 3867..3939
/note="predicted exon, graill2exons human_1.3; frame=0,
forward strand, quality=excellent; putative"
3868..3936
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
...; putative"
exon 4509..4585
/note="predicted exon (exon 1): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=1
misc_feature 4509..4585
/note="predicted exon, graill2exons human_1.3; frame=2,
forward strand, quality=good; putative"
4657..4726
/note="predicted exon (exon 2): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=2
misc_feature 4657..4726
/note="predicted exon, graill2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"
5295..5354
/note="predicted exon (exon 3): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=3
misc_feature 5295..5354
/note="predicted exon, graill2exons human_1.3; frame=2,
forward strand, quality=good; putative"
5441..5577
/note="predicted exon (exon 4): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=4
misc_feature 5441..5577
/note="predicted exon, graill2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"

exon 5714..5798
/note="predicted exon (exon 5): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=5
exon 5915..6009
/note="predicted exon (exon 6): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=6
misc_feature 5921..6009
/note="predicted exon, graill2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"
6093..6216
/note="predicted exon (exon 7): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=7
misc_feature 6093..6216
/note="predicted exon, graill2exons human_1.3; frame=0,
forward strand, quality=good; putative"
/complement(9384..22397)
gene
CDS
/gene="ATP4A"
complement(join(9384..9412,9603..9694,9779..9880,
10216..10349,11806..11951,12475..12598,13691..13845,
13935..14103,14209..14359,14445..14581,15681..15856,
16423..16615,17130..17264,17345..17454,17761..17959,
18573..18841,19131..19383,19587..19700,21203..21406,
21977..22036,22152..22295,22386..22397))
/gene="ATP4A"
/function="human gastric H,K-ATPase catalytic subunit"
/note="human gastric H,K-ATPase catalytic subunit gene
(ATP4A), inverse strand; putative"
/codon_start=1
/product="human gastric H,K-ATPase catalytic subunit"
/protein_id="AAB50172.1"
/db_xref="GI:1905895"
/translation="MGKANEIVLSVVEIPGPGGMAAKMKKKAGGGGKREKLE
MKKMEETNDQQLSVABEOKYQTSATGSLASALLLDVGNPALPBPCTPGYY
PARQLAGLQCLMVAALICLAFATIGSEGLTDNDULIALTAIVAVVPTGCGYY
QEFKSTNIIASFKNLVPQATVIRDGFQFINADQLVVDGLVEMGGRVPADIRILA
AOGKCVNDSLTGSESPQTRSPECHESPLERNTAFSTMCLEGTGVLVYNTDRT
IIGRIASLASGVENKEPTIAIEHFVUIIGALILFATFPIVAMCIGYFLRAMVF
FMATVAVPBGILATVTVCSLTAKRLASNCVKNLEAVETLGSTVICSDKGT
TONRRTVSHLNFPDNIHTADTTEDSQGTPOOSSTETMALCRVLTCRRAPKSGDA
VVPKRIATIGDSSTALKESELTIGNMGGRDRPKVCEPFTNFKPQLSITLLED
PRDPRLHLYMKGAPERVIERCSILIKQELPLDQWMEARQTALSLGGLERVLGF
COLYLNEDYPPGVAFDVEANFPSSGCPFGLVSMIDPPRATVDAYLCKRTAGIRV
IMVTGDHPTAKAIAASVGIISGSEGTIRVDAARLRVVDQVNRDADACVINGQOLK
DMDSSELVEALRTHEWVFARTSPQKLVIVBSCORLGAIVAVTGDGVNDSPLAKKAD

Query Match 100.0%; Score 25; DB 8; Length 40899;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACCAATCT 25
Db 25504 AAAAAAAAAATCGCAACCAATCT 25480

RESULT 2
AC104686/c 149964 bp DNA linear PRI 13-MAY-2005
LOCUS AC104686 BAC clone RP11-74M11 from 4, complete sequence.
DEFINITION Homo sapiens
ACCESSION AC104686 AC024968
VERSION AC104686.2 GI:18855136
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 149964)
AUTHORS Boyer E., Haakenson W. and Radionenko M.
TITLE The sequence of Homo sapiens BAC clone RP11-74M11
JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 149964)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-DEC-2001) Genome Sequencing Center, Washington University, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 149964)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 4 (bases 1 to 149964)
 AUTHORS Wilson, R.K.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAY-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT On Feb 21, 2002 this sequence version replaced gi:17933860.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0074M11
 Drafting Center: WIBR

 NOTICE:
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

 MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

 SOURCE INFORMATION:
 The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
 VECTOR: pBACe3.6

 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is AC005699, 2000 bp overlap: the clone sequenced to the right is RP11-16817. Actual end of this clone is at base position 149964 of RP11-74M11.
 Data from AC024335 was used to finish this clone, AC024968.
 The sequence of AC024968 has been incorporated into AC104686.

 FEATURES
 source
 1..149964
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="RP11-74M11"
 /clone_lib="RPCT-11"

 ORIGIN
 Query Match 100.0%; Score 25; DB 8; Length 149964;

Best Local Similarity 100.0%; Pred. No. 60;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 AAAAAAAAAATCGCAACCAATCT 25
 Db 123240 AAAAAAAAAATCGCAACCAATCT 123216

 RESULT 3
 LOCUS AL359186
 DEFINITION Human DNA sequence from clone RP11-239B6 on chromosome 1 Contains the 3' end of the EIF2C4 gene for eukaryotic translation initiation factor 2C, 4, complete sequence.
 ACCESSION AL359186
 VERSION AL359186.15 GI:21955470
 KEYWORDS HTG; EIF2C4; Eukaryotic translation initiation factor 2C.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 49613)
 White, S.
 Direct Submission
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
 Clone requests: clonerequests@sanger.ac.uk
 On Jul 25, 2002 this sequence version replaced gi:12539691.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; SW: SWSPPROT; Tr: TRMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1
 RP11-239B6 is from the library RPCT-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6

 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: vegas@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

 FEATURES
 source
 1..49613
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-239B6"
 /clone_lib="RPCT-11.1"
 2000
 /note="Clone right end: RP11-435D7"
 join(29424..29501,37908..38009,38605..38946)
 /gene="EIF2C4"
 /locus_tag="RP11-239B6.1-003"
 join(29424..29501,37908..38009,38605..38946)
 /gene="EIF2C4"
 /locus_tag="RP11-239B6.1-003"
 /product="eukaryotic translation initiation factor 2C, 4"
 /note="match: ESTs: Em:CA449759.1"

```
misc_feature
30665..30684
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-003"
/note="Single clone region. Sequence from reads from a
short insert library derived from a single puc clone.
Restriction digest data confirm the assembly."
join(41239..41453,42428..44509)
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-002"
join(41239..41453,42428..44509)
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-002"
/product="eukaryotic translation initiation factor 2C, 4"
/note="match: ESTs: Em:AW295680.1"
join(complement(ALJ54864.16:5815..6078),4633..4798,
10613..10733,13064..13245,13430..13566,13677..13811,
14499..14586,19178..19325,19563..19682,19808..19936,
20188..20321,21741..21925,23585..23744,28916..29231,
29367..29501,37908..38009,38605..38804,41248..45641)
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-001"
join(complement(ALJ54864.16:5815..6078),4633..4798,
10613..10733,13064..13245,13430..13566,13677..13811,
14499..14586,19178..19325,19563..19682,19808..19936,
20188..20321,21741..21925,23585..23744,28916..29231,
29367..29501,37908..38009,38605..38804,41248..45641)
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-001"
/product="eukaryotic translation initiation factor 2C, 4"
/note="match: ESTs: Em:AA482095.1 Em:AA651785.1
Em:A1022110.1 Em:A1674879.1 Em:BB979523.1 Em:BQ214750.1
Em:BQ777272.1 Em:BX103141.1
match: cDNAs: Em:AB046787.1 Em:AB081474.1 Em:AK000040.1
Em:AK030018.1 Em:AK032475.1 Em:AL831908.1"
join(complement(ALJ54864.16:5815..5833),4633..4798,
10613..10733,13064..13245,13430..13566,13677..13811,
14499..14586,19178..19325,19563..19682,19808..19936,
20188..20321,21741..21925,23585..23744,28916..29231,
29367..29501,37908..38009,38605..38804,41248..41356)
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-002"
/standard_name="OTTHUMP000000004431"
/note="match: proteins: Sw:Q5UL18 Tr:Q8BTF4 Tr:Q8CJF8
Tr:Q9HCK5"
/codon_start=1
/product="eukaryotic translation initiation factor 2C, 4"
/protein_id="CAH71584.1"
/db_xref="GI:55664190"
/db_xref="InterPro:IPR003100"
/translation="MEALGPGPASPALFOPPRRGLGTGKPIRLANHFVOVLPKIDV
YHVDVIEKRRPRRREVDPMVHFGQJEGDRPGYDGRMYTAHPIDGR
VUMEVTLRGEQDFPKVSVSVSLQLLEALAGHANEVDDSVQALDVTIRLP
SMRTTPVGRSFFSPBEGYTHPLGGGEVFGFHQSVRPMMMLMIDVSAFTRAQ
PLTEFWCEVLDIQINIEQTKPLTDSQVRFTEIRLKVETHCGQMKRYVCNTR
RASHQTPLEQENGQAMECTVAOYFKQKYSQLQKPHNPLCLOVGEQGHYLPLEVC
NVAGORCICKLTDNOTSTMIKATASAPDROEIKSLVYKSNMGVPPDYKLEFIV
VNMETELTGRVLPAPMLQYGRNKTVAIPNGCVMDMRKOPAFAGEIIVMVAACPA
OKOCREDILKSLFDOLRKT SKDAGMPTGOGPCYKCAQDASEPMEFKLMTYVGLQ
LIVVILPGKTPYAEVAVKVGDTLLGMAITOCVOYKRVKTSPTLSNLCLIKNAKLSGI
NNVLVHORPSVQOQPVIFLADVTHPRPAGDGKPSIAVAVSDMDSHPKRYCATYAVQ
TSROEISOELISQEVQIDLTMMVRELLIQFYKSTFKPTRIIVYGVSEBQMKOVA
WPELAIIRACISILEDYRPGITLYIVQGRHHTRLPCADKTRVGSQVNPAGTYVDS
TTHPEPRFVLCSHAGIQGTSRSHYQVLMQNDCTADELDLITVQLCHTVVRCRS
VGIIPARAYARLVAFPARHYLVKXKDSAEGSVSGSNGRPRQALAKAVQIHDTQH
TMYFA"
polyA_signal
45620..45625
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-001"
45641
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-001"
47614
misc_feature
```

```
ORIGIN
/Note="Clone_left_end: RP4-789D17"

Query Match
Best Local Similarity 93.6%; Score 23.4; DB 8; Length 49613;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGAACAAATCT 25
Db 34925 AAAAAAAAAATCACAACCAATCT 34949

RESULT 4
AC005179/c 59765 bp DNA linear PRI 26-JUN-1998
LOCUS Homo sapiens chromosome 5, p1 clone 1041F10 (LBNL H88), complete
DEFINITION sequence.
AC005179 AC001127 AC001128 AC001129 AC001130 AC001131 AC001578
AC001579 AC002267
VERSION AC005179.1 GI:3258607
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.
1 (bases 1 to 59765)
Kimberly W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
Kadner K., Miguel T., Miller C., Plickuck S., Pollard M.,
Rojeski H., Subramanian S. and Martin C.H.
Sequencing of human chromosome 5
2' (bases 1 to 59765)
Ricke D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
3 (bases 1 to 59765)
Kimberly W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
Davis C.A., Kadner K., Miguel T., Plickuck S., Pollard M.,
Rojeski H., Subramanian S. and Martin C.H.
Direct Submission
Submitted (26-JUN-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
Sequence submitted by:
DOE Joint Genome Institute.
LOCATION/Qualifiers
1..59765
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q"
/clone="P1 1041F10"
/LBNL H88"
56..354
/Note="LBNL H88"
repeat_region
/rpt_family="Alu"
complement(join(882..1069,25942..25994))
/note="Genscan prediction 1"
prim_transcript
1312..1635
/rpt_family="Alu"
complement(join(1187..2055,22602..22663))
/note="81% & 87% identity AL022345"
misc_feature
complement(2422..2661)
/rpt_family="Alu"
complement(2991..3286)
/rpt_family="Alu"
complement(3865..4174)
/rpt_family="Alu"
3872..3893
/note="(T)22"
/rpt_type=tandem
```

```
/rpt_unit="c"
repeat_region 4178..4215
/note="GCT19"
/rpt_type=tandem
/rpt_unit="gt"
repeat_region complement(4240..4540)
/rpt_family="MER42"
repeat_region complement(4739..4803)
/rpt_family="MER42"
repeat_region complement(4865..5216)
/rpt_family="Alu"
repeat_region complement(5288..5407)
/rpt_family="MER20"
prn_transcript join(5861..6093,6630..7108)
/note="99% identity BST dbEST:AA251303 & AA251631"
/prn_transcript complement(5861..6139)
/standard_name="z594906"
repeat_region /note="100% identity EST dbEST:AA416991 & AA470056"
7364..7693
/rpt_family="Alu"
repeat_region 7764..8072
/rpt_family="Alu"
repeat_region complement(8156..8469)
/rpt_family="Alu"
repeat_region 8496..8861
/rpt_family="Alu"
repeat_region complement(8858..9027)
/rpt_family="Alu"
misc_feature 9051..9258
/note="98% identity BST YQ02C04.s1"
/db_xref=dbEST:R89075"
9450..9843
/note="96% identity BST56708"
/db_xref=dbEST:T33112"
9742..10037
/rpt_family="Alu"
prn_transcript complement(10540..11010)
/standard_name="z580610"
repeat_region /note="94% & 98% identity dbEST:AA424065"
10798..10843
/note="(AC)23"
/rpt_type=tandem
/rpt_unit="ac"
repeat_region complement(10872..11053)
/rpt_family="MER46"
repeat_region complement(11904..12023)
/rpt_family="L1"
repeat_region 12279..12775
/rpt_family="MLT2B2"
repeat_region complement(13756..14122)
/rpt_family="MER4"
repeat_region complement(14564..14767)
/rpt_family="Alu"
repeat_region 14761..14822
/rpt_family="L1"
repeat_region complement(14827..15152)
/rpt_family="Alu"
repeat_region 15326..15463
/rpt_family="L1"
repeat_region 16193..16372
/rpt_family="MST"
repeat_region 16467..16580
/rpt_family="MST"
repeat_region complement(17170..17478)
/rpt_family="Alu"
repeat_region 18021..18365
/rpt_family="Alu"
repeat_region complement(18437..18736)
/rpt_family="Alu"
repeat_region 19075..19192
/rpt_family="MER42"
19234..19539
```

```
/rpt_family="Alu"
repeat_region complement(19793..20087)
/rpt_family="Alu"
repeat_region 20568..20718
/rpt_family="Alu"
misc_feature complement(21091..21368)
/note="98% identity EST z139h02.s1"
/db_xref=dbEST:AA676208"
repeat_region 22529..22866
/rpt_family="Alu"
repeat_region 22832..22852
/note="(A)21"
/rpt_type=tandem
/rpt_unit="a"
repeat_region complement(23446..23659)
/rpt_family="MER58A"
repeat_region 24216..24546
/rpt_family="Alu"
repeat_region 24505..24524
/note="(A)20"
/rpt_type=tandem
/rpt_unit="a"
repeat_region 24548..24787
/rpt_family="L1"
repeat_region 25966..26285
/rpt_family="Alu"
repeat_region complement(26552..26847)
/rpt_family="Alu"
repeat_region complement(27205..27504)
/rpt_family="Alu"
repeat_region complement(27522..27746)
/rpt_family="MER30"
28653..28806
/rpt_family="MIR"
29240..29607
/note="98% identity B36828 BAC end sequence"
30284..30374
/rpt_family="LINE2"
32813..33177
/standard_name="endogenous retrovirus"
33453..33574
/rpt_family="Alu"
repeat_region 33661..33829
/rpt_family="Alu"
repeat_region 34077..34716
/rpt_family="Alu"
repeat_region 35152..35345
/rpt_family="MLT1"
repeat_region complement(35525..35660)
/rpt_family="Alu"
repeat_region complement(35726..35889)
/rpt_family="Alu"
repeat_region 35958..36290
/rpt_family="Alu"
misc_feature complement(37374..37494)
/note="GRAT1.2 excellent exon, frame 2"
complement(38108..38761)
/rpt_family="Alu"
repeat_region 38772..39090
/rpt_family="Alu"
repeat_region 39065..39085
/note="(A)21"
/rpt_type=tandem
/rpt_unit="a"
repeat_region complement(39706..39996)
/rpt_family="Alu"
repeat_region 40686..41001
/rpt_family="Alu"
```

Query Match 93.6%; Score 23.4; DB 8; Length 59765;
Best Local Similarity 96.0%; Pred.No. 2.6e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAAACT 25
Db 27642 AAAAAAAAAATCGCAAACT 27618

RESULT 5
AL139234

LOCUS DEFINITION
AL139234 90007 bp DNA linear PRI 18-MAY-2005
Human DNA sequence from clone RP3-438D16 on chromosome Xq24-26.1
contains the PDC8 gene for programmed cell death 8
(apoptosis-inducing factor) (AIF), the RAB33A gene for RAB33A,
member RAS oncogene family (SIO, Rabs10), the 3' end of a novel
gene (FLJ20095) and two Cpg islands, complete sequence.

ACCESSION
AL139234
AL139234.19 GI:14787430
HTG: AIF; Cpg island; FLJ20095; PDC8; RAB33A; Rabs10; SIO.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 90007)
Bird.C.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jul 17, 2001 this sequence version replaced GI:13751266.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChX
RP3-438D16 is from the library RPCT-3 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

FEATURES
source
1..90007
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="q24-26.1"
/clone="RP3-438D16"
/clone_1lb="RPCT-3"
complement(join(7540..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,
16771..16879,17973..18049,18711..18795,23657..23747,
25671..25801,25930..26054,27647..27746,33334..33464,
43728..43929))
/gene="PDC8"
/locus_tag="RP3-438D16.2-002"
complement(join(7540..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,
16771..16879,17973..18049,18711..18795,23657..23747,
25671..25801,25930..26054,27647..27746,33334..33464,
43728..43929))

43728..43929))
/gene="PDC8"
/locus_tag="RP3-438D16.2-002"
/product="programmed cell death 8 (apoptosis-inducing factor)"
/note="match: ESTs: BM927047.1 BY250260.1 BY263326.1
BY281562.1 H15605.1
match: CDNA: AF262320.1 AL049703.1"
complement(join(7540..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,
16771..16879,17973..18049,18711..18795,23657..23747,
25671..25801,25930..26054,27647..27746,33463..33780,
43728..43900))
/gene="PDC8"
/locus_tag="RP3-438D16.2-001"
complement(join(7540..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,
16771..16879,17973..18049,18711..18795,23657..23747,
25671..25801,25930..26054,27647..27746,33463..33780,
43728..43900))
/gene="PDC8"
/locus_tag="RP3-438D16.2-001"
/product="programmed cell death 8 (apoptosis-inducing factor)"
/note="match: ESTs: AL543045.1 AL544137.1 AL545607.1
AL550236.1 AL550264.1 AL557331.1 AL571252.1 AM518690.1
BE887067.1 BE903161.1 BE906258.1 BE027150.1 BF038249.1
BF182905.1 BG108121.1 BG115219.1 BG115901.1 BG116918.1
BG119647.1 BG437720.1 BG744018.1 BG744689.1 B1762702.1
B1826051.1 B1835393.1 BM013834.1 BM015333.1 BM471177.1
BM548814.1 BM807180.1 BM809802.1 EM921817.1 E0058888.1
BQ063233.1 BQ065495.1 BQ066492.1 BQ212050.1 BQ300813.1
BQ308753.1 BQ581468.1 BQ161623.1 BQ165860.1 BQ170548.1
BQ175593.1 BQ517321.1 BQ552706.1 BX331768.1 BX341156.1
match: CDNA: AF100928.1 AF375656.1 AF529274.1 AK000775.1"
complement(join(7540..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,
43728..43857))
/gene="PDC8"
/locus_tag="RP3-438D16.2-003"
complement(join(7540..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,
43728..43857))
/gene="PDC8"
/locus_tag="RP3-438D16.2-003"
/product="programmed cell death 8 (apoptosis-inducing factor)"
/note="match: CDNA: AL049704.1"
complement(7540)
/gene="PDC8"
/locus_tag="RP3-438D16.2-003"
complement(7561..7566)
/gene="PDC8"
/locus_tag="RP3-438D16.2-003"
complement(join(7735..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,
43728..43833))
/gene="PDC8"
/locus_tag="RP3-438D16.2-003"
/standard_name="ORTTHMP00000024009"
/codon_start=1
/product="programmed cell death 8 (apoptosis-inducing factor)"
/protein_id="CA142779.1"
/db_xref="GI:57208843"
/translation="MFRGGGLAAGLKQKLVPTVCVRSROBNRLPARALGNEVI
QLFBEKMGKILPEYLSNMTWEKRRSGVKVMPNATVOSVSSGKLITKIDGRKV
ETDIIVAAGLEPNVELAKTGLTIDSFQGGVAAEIQARSNIVVAGAACFYDIKL
GRRVEVHDHVAVSGRLAGENWTGAELVHQSMMWSLDGPDVGEALIGLVSSLEPTL
GVFAATADQNPKSATGSGTIRSESTESASSETIPSPRAVPAQAPVQCEIDYCKG
VIFVARDVGVIVIMNIENRPIRKIKDGEHEDNEVAKLFNIHED"
complement(join(7735..7806,8148..8344,9853..9977,
11491..11633,14223..14363,14821..14909,15256..15363,

```

16771. .16679,17973. .18049,18711. .18795,23657. .23747,
25671. .25801,25930. .26054,27647. .27746,33334. .33464,
43788. .43833))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
/standard_name="OTTHUMP0000024008"
/codon_start=1
/product="programmed cell death 8 (apoptosis-inducing
factor)"
/protein_id="CA142780.1"
/db_xref="GI:57208844"
/translation="MPRCGGIAGALNKKLVPLVTVCVASPRGRNLPIVQSHHGLSS
PSRSLASGAGKDGSNLYVFLIVAGTVAGVYAATKMEDEKRNERSISGILGTP
EOKRKAALSASGESEEVPODKAPSHVFLILGGTAAPARSIRADPGARVLYSE
DPBLPYMRPPLSKELMTSDDPNVTKTLRFQMKNEKRSIYFGPSPVADQDPLEN
GSAVLVTKGVYQVLDVPRDNWKLNDGSOITREKTLITGCTPRSLAIDAGAKSR
TTLPRKIGDRPSLEKISREKVSITITIGGFGSLALCALGKAKALATVYQIPEPKG
NMGILPPEYLSNMTEKVRREGRVMPNPAIVQSVGSKGLIKLKGREYEDHIVAA
AVGLEPNELAKTGLGLEIDSDPGCFRVAELQASNLVWGDACFPDILKGRVRSH
HDAHVSGRLAGENMTGAAPYMHQSMSESLGCDVCEYALGVDSESLPTVGEAKAT
AODPKSATEOSGTGIRSESETESEASEITTPSTPAVPOAPVQGEDYKGVIFLYLD
KVYVGIYLMINFMNPIARKIIKDGBHEHEDNVAKLFNHED"
complement [join(7735. .7806,8148. .8344,9853. .9977,
11441. .11633,14223. .14363,14821. .14909,15255. .15363,
16711. .16879,17973. .18049,18711. .18795,23657. .23747,
25671. .25801,25930. .26054,27647. .27746,33638. .33780,
43788. .43833))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
/standard_name="OTTHUMP0000024007"
/note="match: proteins: 095831"
/codon_start=1
/product="programmed cell death 8 (apoptosis-inducing
factor)"
/protein_id="CA142778.1"
/db_xref="GI:57208842"
/db_xref="InterPro:IPR001100"
/db_xref="InterPro:IPR001327"
/translation="MPRCGGIAGALNKKLVPLVTVCVASPRGRNLPGNLFORHIV
PLEOTROMASGASGGKIDNSVLVILGSLTVGAGVAYKTMKEDEKRNERSISGL
GLPEOKOKKRAALSASGESEVPODKAPSHVFLILGGTAAPARSIRADPGARVLY
IYSDGPELPYMRPPLSKELMTSDDPNVTKTLRFQMKNEKRSIYFGPSPVADQDPL
HISGVPVLVTKGVQVLDVPRDNWKLNDGSOITREKTLITGCTPRSLAIDAGAK
VKSSTTLFRKIGDRPSLEKISREKVSITITIGGFGSLALCALGKAKALATVYQIPE
PEKNMGKILPPEYLSNMTEKVRREGRVMPNPAIVQSVGSKGLIKLKGREYED
HIVAAVGLPPEYLSNMTEKVRREGRVMPNPAIVQSVGSKGLIKLKGREYED
HIVAAVGLPPEYLSNMTEKVRREGRVMPNPAIVQSVGSKGLIKLKGREYED
RVEHDAHVSGRLAGENMTGAAPYMHQSMSESLGCDVCEYALGVDSESLPTVGEA
AKATDPKSAATGOSGTGIRSESETESEASEITTPSTPAVPOAPVQGEDYKGVIF
YLRKVVVGIYLMINFMNPIARKIIKDGBHEHEDNVAKLFNHED"
complement [join(14876. .14909,15256. .15896)]
/gene="PDCD8"
/locus_tag="RP3-438D16.2-004"
complement [join(14876. .14909,15256. .15896)]
/gene="PDCD8"
/locus_tag="RP3-438D16.2-004"
/product="programmed cell death 8 (apoptosis-inducing
factor)"
/note="match: ESTs: AI834231.1 BC675614.1 BG985116.1"
49826. .63047
/gene="RAB33A"
/locus_tag="RP3-438D16.1-001"

```

```

LOCUS       CINSOLRGR               108736 bp    DNA    linear    PRI 26-APR-2001
DEFINITION  Human chromosome 14 DNA sequence BAC C-2282E14 of library Caltech-D
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL159179
VERSION     AL159179.3   GI:11611160
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
            Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Euteleostomi; Euteleostomi; Primates; Catarrhini;
            Hominoidea; Homo.
REFERENCE   1 (bases 1 to 108736)
            Helliö,R., Pettit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brothier,P., Cattoilco,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBardine,V., Crund,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
            Sequencing of the human chromosome 14
            Unpublished
            2 (bases 1 to 108736)
            Genoscope.
            Direct Submission
            Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - web : www.genoscope.cns.fr)
            On Dec 9, 2000 this sequence version replaced gi:7340700.
            ----- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Genoscope
            Web site: http://www.genoscope.cns.fr/
            Contact: Seqref@genoscope.cns.fr
            -----
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-862P13
            Downstream BAC (overlapping the SP6 end) : R-125H8 (AC=AL135978)
            ----- Summary Statistics
            Assembly program: Phrap, version 2.0
            Quality coverage: 10.09x in Q20 bases, sum-of-contigs
            -----
            Overall quality chart :
            Range      :      bases
            0          :
            1 - 9     :      1
            10 - 19    :      9
            20 - 29    :      40
            30 - 39    :      215
            40 - 49    :      2438
            50 - 59    :      5878
            60 - 69    :      4679
            70 - 79    :      8548
            80 - 89    :      22176
            90 - 99    :      64752
            -----
            Percentage of bases with a quality value >= 40 : 99 %.
            Location/Qualifiers
                source          1..108736
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="14"
                    /clone="C-2282E14"
                    /clone_11b="Caltech-D"
FEATURES             (1)
ORIGIN
Query Match      93.6%; Score 23.4; DB 8; Length 108736;
Best Local Similarity 96.0%; Pred. NO. 2.1e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Gy      1 AAAAAAAAAATGCAACAACTCT 25
Gb      69597 AAAAAAAAAATGCAACAACTCT 69573

```

RESULT 7
BX842242/c 111002 bp DNA linear PRI 17-APR-2005
LOCUS Homo sapiens chromosome 10 clone RP11-436019, partial sequence.
DEFINITION BX842242 AC061711
ACCESSION BX842242.1 GI:38489886
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 111002)
Whitehead, S.
REFERENCE Direct Submission
Submitted (19-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
On Nov 21, 2003 this sequence version replaced gi:20564422.
This sequence was originally produced by Genome Therapeutics
Corporation 100 Beaver Street, Waltham, MA 02453, USA.
The clone was found to contain a deletion, and has been resubmitted
in two pieces to aid construction of the chromosome 10 tiling path.
FEATURES
source
1..111002 Homo sapiens
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RP11-436019"
/clone_1ib="RP11-11.2"
ORIGIN
Query Match 93.6%; Score 23.4; DB 8; Length 111002;
Best Local Similarity 96.0%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 80277 AAAAAAAAAATCGCAACAAATCT 80253
RESULT 8
AL139143 111862 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP4-555P23 map p34.1-34.3, 16
DEFINITION unordered pieces.
ACCESSION AL139143
VERSION AL139143.6 GI:9863524
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1
Plumb, B.
REFERENCE Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Request: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9212183.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj555P23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; l08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 105309 bases at least Q40
Consensus quality: 107835 bases at least Q30
Consensus quality: 109154 bases at least Q20
Insert size: 110362; sum-of-contigs
Insert size: 138096; 6.8% error; agarose-fp
Quality coverage: 3.57x in Q20 bases; sum-of-contigs Quality
coverage: 3.11x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6111 6110: contig of 6110 bp in length
6211 6210: gap of 100 bp
12365 12365: contig of 6135 bp in length
12366 12465: gap of 100 bp
12466 14908: contig of 2443 bp in length
14909 15008: gap of 100 bp
15009 17724: contig of 2716 bp in length
17725 17824: gap of 100 bp
17825 25035: contig of 7211 bp in length
25036 25135: gap of 100 bp
25136 27936: contig of 2801 bp in length
27937 28036: gap of 100 bp
28037 31938: contig of 3902 bp in length
31939 32038: gap of 100 bp
32039 44844: contig of 12806 bp in length
44845 44944: gap of 100 bp
44945 57247: contig of 12303 bp in length
57248 57347: gap of 100 bp
57348 63813: contig of 6466 bp in length
63814 63913: gap of 100 bp
63914 70362: contig of 6449 bp in length
70363 70462: gap of 100 bp
70463 80246: contig of 9784 bp in length
80247 80346: gap of 100 bp
80347 94000: contig of 13654 bp in length
94001 94100: gap of 100 bp
94101 101709: contig of 7609 bp in length
101710 101809: gap of 100 bp
101810 107002: contig of 5193 bp in length
107003 107102: gap of 100 bp
107103 111862: contig of 4760 bp in length.
FEATURES
source
1..111862 Homo sapiens
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p34.1-34.3"
/clone="RP4-555P23"
/clone_1ib="RP11-4"
1..6110
/note="assembly fragment: 00533
fragment chain:1"
6211..12365
/note="assembly fragment: 00488
fragment chain:1"
12466..14908
/note="assembly fragment: 00991
fragment chain:1"
15009..17724
/note="assembly fragment: 00166
fragment chain:2"
17825..25035
/note="assembly fragment: 00257
fragment chain:2"
25136..27936
/note="assembly fragment: 00608
fragment chain:2"

Fragment chain:3"
misc_feature 28037..31938
/note="assembly_fragment:00316
fragment_chain:3"
misc_feature 32039..44844
/note="assembly_fragment:00611
fragment_chain:4"
misc_feature 44945..57247
/note="assembly_fragment:00630
fragment_chain:4"
misc_feature 57348..63813
/note="assembly_fragment:00206.0"
misc_feature 63914..70362
/note="assembly_fragment:00382"
misc_feature 70463..80246
/note="assembly_fragment:00429"
misc_feature 80347..94000
/note="assembly_fragment:00431"
misc_feature 94101..101709
/note="assembly_fragment:00467"
misc_feature 101810..107002
/note="assembly_fragment:00863"
misc_feature 107103..111862
/note="assembly_fragment:00880"

Query Match 93.6%; Score 23.4; DB 14; Length 111862;
Best Local Similarity 96.0%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25
Db 77558 AAAAAAAAAATCACAACAATCT 77582

RESULT 9 AC025781 112862 bp DNA linear PLN 19-JAN-2001
LOCUS AC025781
DEFINITION Arabidopsis thaliana chromosome 1 BAC F15C21 genomic sequence,
complete sequence.
ACCESSION AC025781
VERSION AC025781.8 GI:12322475
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Bukeriyota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosoids; eurosoids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 112862)
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maitl,R., Renning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Frazer,C.M.
Arabidopsis thaliana chromosome 1 BAC F15C21 genomic sequence
Unpublished
2 (bases 1 to 112862)
Lin,X. and Kaul,S.
Direct Submission
Submitted (14-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 112862)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced g1:12280766.
Addres all correspondence to:at@tigr.org

BAC clone F15C21 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from Sp6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including GenScan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GenemarkMM (Mark Borodovsky,

http://genemark.biology.gatech.edu/Genemark/), Glimmer4 (a variant
of Glimmer3, see Mihalea Perlea,
http://www.tigr.org/softlab/glimmer4/glimmer4.html, and
GenesSplicer (Mihalea Perlea and Steven Salzberg, contact
mperlea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES
source

gene
CDS
mRNA
/gene="F15C21.1"
/note="similar to GI:1100253 from [Arabidopsis thaliana]"
join(<1..24,120..307,390..440,515..554,643..767,844..960,
1036..1110,1187..1243,1361..1465,1577..1663,1761..1952,
2027..2215,2295..2405,2505..2594,2677..2853,2944..3021,
3097..3219,3293..3562,3666..3836,3858..3997,4077..4314,
4395..4747,4828..5041,5137..5313,5401..7551,7654..7871)
/gene="F15C21.1"
join(<1..24,120..307,390..440,515..554,643..767,844..960,
1036..1110,1187..1243,1361..1465,1577..1663,1761..1952,
2027..2215,2295..2405,2505..2594,2677..2853,2944..3021,
3097..3219,3293..3562,3666..3836,3858..3997,4077..4314,
4395..4747,4828..5041,5137..5313,5401..7551,7654..7871)
/product="acetyl-CoA carboxylase, putative, 5' partial;
1-7710"
/protein_id="AAGS1250.1"
/db_xref="GI:12322477"
/translation="TTPMQIPETIRFYGIEHGQYDWMKRTSVVAFPPEDKQSIIRP
KGHCVAIVRVSSEDDDFKPTSGVGLSEFKSNVWAFVSKGSGIHFRSDQFGH
VFAGSEBALAIANMVLGLKEIOIRGETRTVDYITDILHSDYRDNKHITGMDSRI
AMRVARPRPMPVLSVGVGLYKASAAVSDVGLKGOIPPKHSIVHSGVSN
IEGKTYTIDVRRGSGTTRLRNKSSEVVAEHTTRDGLMLQDGKSHVITAEBAAG
TRLIDRTCLQNDHDPSSKMASTPCKMLKYLISDNSNIDATPVAVEWMKCMPL
LSPSGVTHFPMSSGQMOGELIANLDLDPSSVRAEPHGFPRGLPTAISGRV
HORCAATLNARMIAGYEYKVDVVDLNCDSPELPEOMGECAVLAITRPLKLN
HMCESKYREPESIRNSLTDPFAKLGLILEHLSGSCDEKGERALREPLKSLA
KSYGGRGSHARVIVHSLEFVSEELPDMNLTADVITRMOIYKDDKIDVIVIS
HOGIKRNGKVLRLMEQIVPNPAAYNDKILRSTLWHTNSELALASOLLEBKTS
ELRSLNARSLSLEMTFDEKNDTPRKSLINRISLDLVSALVADALGEBTCKLS
HTLRNARVETVIRLYOPVYKDSVRQMQHRSGLTAPELEEMERKNGLIDPDS
IEKGLVGRSKRKQWAMFLPSIISALRETKNDYETGAPLSPGMMKIAIYGINQDS
LLOSGDEDOAQRVNLAKLKEKEVSSISGAGVVISCTIORDGRTPMRSPFHM
SLKQVYVEELPLHPLSLIYELKIKGYSNIOYTPSDRQMHLYTVYDKVPLK
RMPRLSVROATNDGFTILOOGDOKLSQTLISAFPSKCYLRSLMAMELEMANH
AAMKPDHMHFLCTRQIDDLVFFRRRVANVAGDEETVEMLEBARITHISVG
RMHRLGCEWMEVRLMVSGLACAMVVAANDECTVAHYIEVETPGNSLIYHS
ITKKGPLHEPIISQYKPLGLDQRLAARSNTYCYDEPLAGTALDELMIAQOHS
VKKYEKDTLINVKLEFSKESGSGSLDIVERPGNDPDMVACIDMSPEPMPOR
KLIVANDVTFKSGSGPRDPAFLATYELCAKCKPLIYLAASGARLVGAEBVKIC
EKVMSDEIPEKNGFYIYSPREHIGSSVLAHYKLSGSETRWITDTYVGEDCI
GVENTSGSAGIAGVSKAYNBTFTLVTSQGTGIGALAKLGRCTQRDQPLITLG
FSLTNKLIGREYVSHMQLGDKPMNGVNLVSDLGGVSAIIMWLSTIPAYVG
PVTNLADLPERRVIEVPSNCDPRAIAGKVDKNTKGLFDKNSFIETLGGAR
TVVTGRKLGGIPIGVVAVEQTVMQIIPADPGQDGHERRVPAQGVWFPDYSAKTA
QALMDFYREELPLILANMGFGSGQGLDPEGLQASSTIVENRTYROPVFIYPMW

GELRGAMVUNDSQINSVDYVEMYADBTAGNVLBEGCTEIEIFRTKELLECGRLDOK
 LLSLAKLQADAKQSEAYANIELLOQITKAREQOLPEVYIQITKPELHDTSMRAAK
 GVIKSVENSGSRSPFFYKKNLRRIASSLVKNVRSGNGLVYKSMRLIQOMFNSD
 IAKGKEAWTDQVFEFTKDNNSYNELKLSIELRAQTLNLQALAEIGNSDLOALPQILA
 NLLNKVPEPSKREELVAARKVIG"

9984 . 22276
 /gene="P15C21.2"
 /note="similar to GI:1100253 from [Arabidopsis thaliana]"
 join(<9984 . 10248,10452 . 10834,10913 . 11119,11200 . 11361,
 11439 . 11531,11617 . 11712,14225 . 14398,14480 . 14573,
 14658 . 14845,14929 . 14979,15063 . 15136,15190 . 15313,
 15395 . 15511,15587 . 15661,15739 . 15795,15921 . 16025,
 16136 . 16222,16320 . 16511,16597 . 16785,16867 . 16977,
 17077 . 17166,17253 . 17429,17519 . 17596,17680 . 17802,
 17877 . 18146,18252 . 18580,18676 . 18743,18765 . 18913,
 19000 . 19352,19437 . 19650,19738 . 19914,20001 . 22151,
 22220 . 22276)
 /gene="P15C21.2"
 /cdoon start=1
 /product="acetyl-CoA carboxylase, putative: 9984-22276"
 /protein_id="AA051252.1"
 /db_xref="GI:12322479"
 /translation="MEMRALGSSCTGNGSAPITLTNTSPWITTFPSTVLRSSLR
 TFKGSSRVRTFKGVSTRTVLSRTKOQFLPFLNDFPLENDVSEKRYTVLPDG
 SYNGASVNGSHSDVVPGRNVAEVENFCALGKQPIHSLIIVATNGVSEKRYTVLP
 MAYETGSEKAYKIVAMATPEMRINABEIRIADQVREYRGCTNNNNVAYIEMA
 EYTRDAVAPRGHASENELPDALKEKITTFLGPPADSMIALGDITGSLIQAQADV
 PTLPSGSHVAKIPGKSLVTPBEIYKKAQVTTTEBALASCOVGPANIKASWGGG
 KQIRKRVHNDIEVALFKQVQGEVPGSPPIFMVYASQSRLEAQLCDQYGNVALSHR
 DCSVORHOKITIEEGPIVAPOETIKLEQAAARLAKSVNVAQAAVEVLEMDTEY
 YTELNPRLOVHEPVTWIAEVNLPAAOVAVMGPIPLMOIPEIRIORYGMEHGGQMS
 RCTSVASPDPEAESLAPKGCVAVRPESDPDQGEIPEIRIORYGMEHGGQMS
 YSVASGGGIEHPDSQFKLNEKRGMLQHPARESSVALIANVTLAKITQIRGD
 IRTNDVYITDLHASYRENKIHTGWLDIRIMRVAEAPRYLVSQVGLAKASTS
 SAVSDYVGLERKQIPRKHISLVHSQVLNIEGSRVTTIDVVRGSGYRLKMSSEV
 VAEIHTLRDGLMLQDKSHVYIAKEEATGTRLLDGTCLQNDHPSKMAEPRC
 KILRYLVNNSSIDTDTPAEYVVMCMPLISPAVGVHPLSEQAOAGELAKL
 DLDPSAVKAKPRGSPRILGPTAISCKVHRCQACTNNAAMLACVDHVDVDTQ
 DILNCLDSPELPLQOECPAYLATRLPDOLNMLELTKKEPEIISKTSITLDPFKL
 LGILLEAHLSSCDKERGSLERLIEPLMSLVSEGGRESHARLIVHSLFEELYVEE
 LFNDMLADVIRMQOYKQKRLKVIDYLSHQIHKRKLVLRLMEQLVYNPAPAYR
 EKLIRFSAINTHTNYSQLAKASOLLEQTRSELSNARSISELEFTAGMDPDK
 RKSASETMENTVSSSLAEDLVGLFHDSDHTRQVVEYTIHRLYQYVQESRPI
 OMHOGSVIASWELREHFERKNGTGDHETSEKGIYVASKRRGRTVITKSIQFESI
 TNASLRETHSHCEYKRAPLSCNMHAIIVGINNMQSLQDSGDEQOTDERNKLAKI
 LKEEVSVGVISCIQIRDEGRTPMRSHFMLEKQYVEEPLRHVEPLSVLELDEK
 LKGSNIQVSPSRQMWNYSVYDRPVRKRWLRSLVROTTMNDGFLLOQSDYOLS
 QTVLSMAFTSKILRSINMABELELNANAAKPDHAMEFLCILEEQIDILVPRP
 REVNADEETIVETILEATOEIHSVGRMHALGCEMVEVIMVSSGLGAPARV
 VYANTVGRCTVHIVREVATGRNLSIYHSITKKGPIHGTILINGOVKPIANNIDRELA
 AARSNTTYCYDPLAFETLLELMASQHSQVKKPCPNRLINKVCELFSTBSLGTSL
 IEVERPAGLNDIGMVAWILMESTPEFPMKRLIVANDYFKASGSPREDAPFLAVT
 ELACTKGLPLIYLAANSARGLVAEVEKACFKVMSGDSPGNDPOYILASEEDVARI
 GSSVIAHEVKLPSEGETRNVITDVGKEDGLGAVENLGSAGIAGYRAVNETVLFV
 SGRSVIGIAYLARIAGRCIORLDOPILILGFTLNLKREYVSSHMOGCPKIMT
 GVYHLTVSDDLRGVSAIINMLQPIIPAYVGLPLVLAPLDPERTVYIPENSCDMPAA
 IAGINDNTGKMLGGITFDKXSPVETTESGARTVYTGAKIGGPIGVANETQVWHTV
 PADPGOLDSHERVPOGOVWEPDSAAKQAQALMWNVRQOLPFIITANRGSGGGRD
 LEFGILOQSAIVENLRTYRQVYVYIPMGELRGAMVNVVQINSIDYIEMVABETA
 RSNVLEPEGMIIEKFRKELIECMGRIDOTLINIKANIDADAKNKAYANIELLOQIK
 TREKOLPYVTOIATKEALHTSMMAKGVKSVENSGSRSPFFYKKNLRRIASS
 LVNIRKASGDLISYKAMGLIQDMFRKSEIAKGEAWTDQVFEFTKDNNSYNELQK
 LSELRTQKLIQNLAEIGNSDLOALPQILANLKNKVDLSRREELVAARKVIG"

repeat_region	13141 . 13243
/rpt_family="ATREP1 ATREP1 ATREP1 dispersed repeat from GB DB6608 - a consensus"	
gene	24061 . 28523
/gene="P15C21.3"	
/note="pseudogene, polyprotein"	
/pseudo	
repeat_region	24350 . 24662
/rpt_family="RE1 RE1 Arabidopsis thaliana retroelement - a consensus"	
repeat_region	25895 . 26657
/rpt_family="ATCOP1A15 ATCOP1A15 Internal region of ATCOP1A15 copia-like LTR-retrotransposon."	
repeat_region	25998 . 26405
/rpt_family="ATCOP1A11 ATCOP1A11 Internal region of ATCOP1A11 LTR-retrotransposon."	
repeat_region	26000 . 26684
/rpt_family="ATCOP1A16 ATCOP1A16 Internal region of	

Query Match	93.6%; Score 23.4; DB 15; Length 112862;
Best Local Similarity	96.0%; Pred. No. 2.1e+02;
Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATCT 25
 DB 104314 AAAAAAAAAATCGCAACCAATCT 104338

RESULT 10	AP001628
LOCUS	AP001628 114835 bp DNA linear PRI 03-JUN-2000
DEFINITION	Hom sapiens genomic DNA, chromosome 21, clone:KB51A8, MW1-D21S171 region, complete sequence.
ACCESSION	AP001628
VERSION	AP001628.1 GI:7670582
KEYWORDS	HTG.
SOURCE	HTG.
ORGANISM	Homo sapiens (human)
REFERENCE	1 (bases 1 to 114835)
AUTHORS	Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE	Hom sapiens genomic DNA, chromosome 21, clone:KB51A8, MW1-D21S171 region
JOURNAL	Published Only in Database (2000)
REFERENCE	2 (bases 1 to 114835)
AUTHORS	Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE	Direct Submission
JOURNAL	Submitted (04-Apr-2000) Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan (E-mail:nshimizu@mb.med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)

FEATURES	location/Qualifiers
source	1..114835
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="21"
	/clone="KB51A8"
	/cell_line="FLEB14-14"
	/cell_type="pre-pro-B cell"
	/clone_1fb="Keio BAC library"

ORIGIN	
Query Match	93.6%; Score 23.4; DB 8; Length 114835;
Best Local Similarity	96.0%; Pred. No. 2.1e+02;
Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATCT 25
 DB 92733 AAAAAAAAAATCGCAACCAATCT 92757

```

RESULT 11
AC01510/c 129402 bp DNA linear PRI 23-OCT-2000
LOCUS Homo sapiens chromosome 19 clone CTD-2195B23, complete sequence.
AC01510
AC01510.7 GI:10947023
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE 1 (bases 1 to 129402)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 129402)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL DOE Joint Genome Institute.
REFERENCE Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL 3 (bases 1 to 129402)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (23-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Oct 23, 2000 this sequence version replaced gi:9211205.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing completed at Stanford Human Genome Center
www.sshc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.7.
FEATURES
source
1..129402
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2195B23"
ORIGIN
Query Match 93.6%; Score 23.4; DB 8; Length 129402;
Best Local Similarity 96.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 1;
Qy 1 AAAAAAAAAATCGCAACCAATCT 25
Db 65509 AAAAAAAAAATCGCAACCAATCT 65485
RESULT 12
AC125612/c 135754 bp DNA linear PRI 13-MAR-2003
LOCUS Homo sapiens 12 BAC RP11-117N2 (Roswell Park Cancer Institute Human
AC125612
AC125612.7 GI:23346643
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 135754)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Albrooks,S.L., Amaralunga,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouch,V., Bowie,S., Brileva,M., Brown,B., Brown,M., Brynild,N.P.,
Burch,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,

```

```

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Daviola,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlik,P.,
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M.,
Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,D., Hulyk,S.,
Hune,J., Iosifides,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Korvan,J., Kovar,C., Kravovic,J., Kuresh,A., Landry,N.,
Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozano,R.J., Lu,X.,
Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
Mashney,E., McLeod,M.P., Meador,M., Mei,G., Mescher,S.,
Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
Monabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okunolu,G.,
Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shim,C., Shooshari,N., Sisson,I.,
Sodergren,B., Sotak,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syarik,A., Tabor,P., Tameris,A., Tameris,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vazquez,L., Vera,V., Villalón,D., Vinson,R., Wang,D.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.P., Zhou,J., Zorrilla,S., Zuchelapati,R.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 135754)
Worley,K.C.
Direct Submission
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 135754)
Worley,K.C.
Direct Submission
Submitted (16-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 135754)
Worley,K.C.
Direct Submission
Submitted (30-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 135754)
Worley,K.C.
Direct Submission
Submitted (13-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 30, 2002 this sequence version replaced gi:22901969.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

FEATURES

source

Location/Qualifiers
1..135754
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-117N2"
/complement(1..3336)
/note="overlaps bases 1..3336 of clone AC021052"
/function="clone overlap"
39..328
/rpt_family="AluSx"
complement(820..1021)
/rpt_family="MIR"
complement(1030..1329)
/rpt_family="AluJo"
1540..1563
/rpt_family="TTTTA)n"
2183..2480
/rpt_family="AluSc"
complement(2723..3091)
/rpt_family="LMEL1"
complement(3144..3452)
/rpt_family="AluSx"
3937..4053
/rpt_family="MLT1K"
4181..4271
/rpt_family="MLT1K"
complement(4272..4570)
/rpt_family="AluSx"
4571..4645
/rpt_family="MLT1K"
complement(4663..4819)
/rpt_family="MIR"
5283..5585
/rpt_family="AluSp"
5595..5904
/rpt_family="AluSp"
5915..6218
/rpt_family="AluSx"
6382..6650
/rpt_family="AluSx"
6654..6690
/rpt_family="(CAAA)n"
complement(7736..7915)
/rpt_family="MERSA"
8106..8268

repeat_region /rpt_family="MER103"
complement(8438..8736)
repeat_region /rpt_family="AluSx"
8779..9032
repeat_region /rpt_family="AluSx"
9035..9063
repeat_region /rpt_family="(TAA)n"
9064..9182
repeat_region /rpt_family="FLAM_C"
9186..9211
repeat_region /rpt_family="(TAA)n"
9277..9581
repeat_region /rpt_family="AluSg"
complement(9626..9753)
repeat_region /rpt_family="AluSg/x"
11338..11381
repeat_region /rpt_family="(CCCCG)n"
11665..11703
repeat_region /rpt_family="GC_rich"
14667..14704
repeat_region /rpt_family="AT_rich"
15189..15216
repeat_region /rpt_family="(TTTG)n"
complement(15985..16362)
repeat_region /rpt_family="MLT1A2"
complement(16405..16566)
repeat_region /rpt_family="MIR"
16779..16803
repeat_region /rpt_family="AT_rich"
16804..17094
repeat_region /rpt_family="AluSc"
17095..17147
repeat_region /rpt_family="AT_rich"
complement(17155..17441)
repeat_region /rpt_family="AluSx"

Query Match 93.6%; Score 23.4; DB 8; Length 135754;
Best Local Similarity 96.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 1;

QY 1 AAAAAAAAAATCGCAACAATCT 25
Db 121951 AAAAAAAAAATCGCAACAATCT 121927

RESULT 13
HSDJ792G4 142728 bp DNA linear PRI 18-MAY-2005
LOCUS HSDJ792G4
DEFINITION Human DNA sequence from clone RP4-792G4 on chromosome 1p31.2-32.2
contains the 5' end of the ALG6 gene for asparagine-linked
glycosylation 6 homolog (Yeast alpha-1-3-glucosyltransferase), a
dual (Hsp40) homolog subfamily C member 7 (DNAJC7) pseudogene, the
FOXN3 gene for forehead box D3, a novel gene, the 5' end of a novel
gene and a CpG island, complete sequence.

ACCESSION AL049636
VERSION AL049636.22 GI:11967799
KEYWORDS HTG; ALG6; CpG island; DNAJC7; FOXN3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 142728)
Howden, P.

REFERENCE Direct Submission
AUTHORS Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
TITLE Clone request: clonerequest@sanger.ac.uk
JOURNAL On Dec 22, 2000 this sequence version replaced gi:5531522.

COMMENT The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMBEP; Information
on the WORMBEP database can be found at

[illegible]

except on the rare occasion of the phone being a fax.

polYA_signal

polYA_site

gene

CDS

gene

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 8; Length 142728;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCCGAAACAATCT 25
|||||
Db 66762 AAAAAAAAAATCCGAAACAATCT 66766

RESULT 14

LOCUS AL137881/c

DEFINITION Human DNA sequence from clone Rpl1-40A8 on chromosome 13 Contains a novel gene, the 3' end of the GUCY1B2 gene for guanylate cyclase 1 soluble beta 2 and a CpG island, complete sequence.

ACCESSION AL137881

VERSION AL137881.12 GI:8894206

KEYWORDS HTG; Cpg Island; guanylate cyclase; GUCY1B2.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 143324)
Pelan,S.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jul 1, 2000 this sequence version replaced gi:8546598.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Mp., MORMPREP; Information on the MORMPREP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormrep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
RPl1-40A8 is from the library RPc1-11.1 constructed by the group of Plier de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

VECTOR: pBACe3.6

Genome Center

```

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
Location/Qualifiers
1.143324
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-40A8"
/clone_1fb="RP11-11.1"
1
/note="Clone_left_end: RP11-40A8"
14112
/note="Clone_right_end: RP11-233H19"
19526.19591
/note="Single clone region. Region contains reads from
clone pcr only (oligos 1&2). Assembly is consistent with
restriction digest."
complement(join(25997..26264,33228..33354,54102..54331))
/locus_tag="RP11-40A8.3-001"
complement(join(25997..26264,33228..33354,54102..54331))
/locus_tag="RP11-40A8.3-001"
/note="match: ESTs: BF574759.1"
join(53410..53759,71026..71097,73094..73201,74302..74378,
78504..78618,86940..87013,89046..89151,91606..91687,
93082..93124,97524..97604,99977..100383)
/locus_tag="RP11-40A8.1-001"
join(53410..53759,71026..71097,73094..73201,74302..74378,
78504..78618,86940..87013,89046..89151,91606..91687,
93082..93124,97524..97604,99977..100383)
/note="match: ESTs: AL520828,AL539901,AL568105,AU19164,
BG259079,BG259492,BG261287,BG281634,BG333104,BG724123,
B1660913,B1754776,BM906357,BM18887,BM999525,BC016444,
BQ216110
match: CDNAs: AK021774,BC001397,BC005088,BC007332,
BC010174"
join(53410..53759,56614..56748,71024..71085)
/locus_tag="RP11-40A8.1-002"
join(53410..53759,56614..56748,71024..71085)
/locus_tag="RP11-40A8.1-002"
/note="match: ESTs: BM847956.1"
join(53422..53524,56612..56748,71024..71097,73094..73201,
74302..74378,86940..87002)
/locus_tag="RP11-40A8.1-005"
join(53422..53524,56612..56748,71024..71097,73094..73201,
74302..74378,86940..87002)
/locus_tag="RP11-40A8.1-005"
/note="match: ESTs: B1753018.1"
join(53696..53759,71026..71097,73094..73201,74302..74378,
78504..78618,86940..87013,89046..89151,91606..91687,
93082..93124,97524..97604,99977..>100090)
/locus_tag="RP11-40A8.1-001"
/standard_name="OTTHUMP00000018436"
/note="match: proteins: Q9D014 Q9HAF7"
/codon_start=1
/protein_id="CA113913.1"
/db_xref="GI:55937387"
/translation="MARGDCGCGVGRARPIVSEVYLKDSKMKNGMLPVLKLNPC
SGEGATYLFMCIQQLFEVAKFEKHSNPITNSVSGGLIHPPTPDPFLFLHYL
KDKKEGFQPLDDVVNDVNEPNCILIKLGLGELHLHVTEKKNPEIDDKRYKYSK
EKTLKLEKKVQNTVAALKTNNNVSNSTFESDDGASTDEEDYIRAVHQLISD

```

```

YIPKELSDLSKYLKLPSPASLPPNPSPKIKLSDEPEYKEDYTKFNTKDLKTEKKN
SKMTAQAQALAKVDSKMSKIDTFRGVSNKKKIGV"
misc_feature
69479
//locus_tag="RP11-40A8.1-002"
//note="Clone Left end: RP11-547C18"
gene
join(71981..72127,73094..73201,74302..74378,78504..78618,
86940..87013,89046..89151,93080..93124,97524..97604)
//locus_tag="RP11-40A8.1-004"
join(71981..72127,73094..73201,74302..74378,78504..78618,
86940..87013,89046..89151,93080..93124,97524..97604)
//locus_tag="RP11-40A8.1-004"
//note="match: ESTs: BM559365.1"
gene
join(73733..73774,74302..74378,86940..87013,89046..89133)
//locus_tag="RP11-40A8.1-003"
join(73733..73774,74302..74378,86940..87013,89046..89133)
//locus_tag="RP11-40A8.1-003"
//note="match: ESTs: BE972426.1"
gene
join(93101..93124,93690..93937,97524..97604,99977..100121)
//locus_tag="RP11-40A8.1-006"
join(93101..93124,93690..93937,97524..97604,99977..100121)
//locus_tag="RP11-40A8.1-006"
//note="match: ESTs: AW43381.1"
polyA_signal
100360..100365
//locus_tag="RP11-40A8.1-001"
100383
polyA_site
//locus_tag="RP11-40A8.1-001"
join(Complement(Al160157.17:63216..63289),
Complement(Al160157.17:48615..48709),
Complement(Al160157.17:34398..34554),
Complement(Al160157.17:29893..30088),
Complement(Al160157.17:28603..28799),
Complement(Al160157.17:27097..27208),
Complement(Al160157.17:25729..25832),
Complement(Al160157.17:24795..24711),
Complement(Al160157.17:20793..21133),
Complement(Al160157.17:17311..17459),
Complement(Al160157.17:16576..16722),
Complement(Al160157.17:10468..10622),
Complement(Al160157.17:7385..7536),
Complement(Al160157.17:7095..7304),
Complement(Al160157.17:4648..4971),
Complement(138130..138430))
//gene="GUCY1B2"
//locus_tag="RP11-40A8.2-001"
join(Complement(Al160157.17:66346..66401),
Complement(Al160157.17:63216..63289),
Complement(Al160157.17:48615..48709),
Complement(Al160157.17:34398..34554),
Complement(Al160157.17:29893..30088),
Complement(Al160157.17:28603..28799),
Complement(Al160157.17:27097..27208),
Complement(Al160157.17:25729..25832),
Complement(Al160157.17:24795..24711),
Complement(Al160157.17:20793..21133),
Complement(Al160157.17:17311..17459),
Complement(Al160157.17:16576..16722),
Complement(Al160157.17:10468..10622),
Complement(Al160157.17:7385..7536),
Complement(Al160157.17:7095..7304),
Complement(Al160157.17:4648..4971),
Complement(138130..138430))
//gene="GUCY1B2"
//locus_tag="RP11-40A8.2-001"
//product="guanylate cyclase 1, soluble, beta 2"
//note="match: ESTs: A1247180 A1792818 A1822009 AW300728
BG181663 BG680126.1
match: CDNA: AF038499"
polyA_site
//gene="GUCY1B2"
//locus_tag="RP11-40A8.2-001"
Complement(138150..138155)
/gene="GUCY1B2"
polyA_signal

```

```

misc_feature
143324
//locus_tag="RP11-40A8.2-001"
143324
//note="Clone_right_end: RP11-40A8"
ORIGIN
Query Match 93.6%; Score 23.4; DB 8; Length 143324;
Best Local Similarity 96.0%; Pred. No. 2e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 23943 AAAAAAAAAATCGCAACAAATCT 23919
RESULT 15
AL136128
LOCUS
DEFINITION
Human DNA sequence from clone RP1-84N20 on chromosome 6 Contains a
novel gene, the 5' end of the TPD52L1 gene encoding two variants of
tumor protein D52-like 1 protein and two CpG islands, complete
sequence.
AL136128 151875 bp DNA linear PRI 18-MAY-2005
AL136128.9 GI:10443357
HTG; CpG island; D52; TPD52L1.
SOURCE
Homo sapiens (human)
Organism
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 151875)
Almeida,J.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Oct 1, 2000 this sequence version replaced gi:9796103.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP1-84N20 is from the library RP1-1 constructed by the group of
Piter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES
Location/Qualifiers
1..151875
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704N2084"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP1-84N20"
/clone_id="RPCI-1"
5621
//note="Clone_right_end: RP11-510H23"
misc_feature

```

gene
join(18678..18917,31415..31579,50130..50331,54909..54980,55922..56036,56337..56988)
/locus_tag="RPl-84N20.1-003"
join(18678..18917,31415..31579,50130..50331,54909..54980,55922..56036,56337..56988)
/locus_tag="RPl-84N20.1-003"
/note="match: cDNAB: BC026087"
join(18678..18917,31415..31579,50130..50331,54909..54980,55922..56036,56337..56370)
/locus_tag="RPl-84N20.1-003"
/standard_name="OTHUMP00000017132"
/codon_start=1
/protein_id="CA122999.1"
/db_xref="GI:56204337"
/db_xref="InterPro:IPR002867"
/translation="MOVOLGOVEIKCPTECEFLPETTVVYNLTDESIKYKYLELGRIDSTKCPCKHPTTPKKKGHIPTPSRSSTKTIQPTPCQFWCKCHSPHMEGVNCKEYKRGGLRLHNSLIEHGQRNAQCPKCKIHIOREGCDDHMTCSQCNTPNFCRGGERKKYRFGDHTNSLIFGCKRYRLPERPHLRILVSGVACGLFIAPLIMVGLALCAIVAIYEELIKTYNLISGRTRMOTQILAQVPLSDMYCLKQVETICISVLLPL"
join(18678..18917,31415..31579,50130..50331)
/locus_tag="RPl-84N20.1-002"
join(18678..18917,31415..31579,50130..50331)
/locus_tag="RPl-84N20.1-002"
/note="match: ESTs: BF045754 BF467827 BF556843 BG025210 BG726272"
join(18678..18917,31415..31579,50130..50331)
/locus_tag="RPl-84N20.1-002"
/standard_name="OTHUMP00000017131"
/note="match: proteins: Q9BX48"
/codon_start=1
/protein_id="CAC36346.1"
/db_xref="GI:13591434"
/db_xref="GeneW:21487"
/db_xref="GOA:Q8TC41"
/translation="MOVOLGOVEIKCPTECEFLPETTVVYNLTDESIKYKYLELGRIDSTKCPCKHPTTPKKKGHIPTPSRSSTKTIQPTPCQFWCKCHSPHMEGVNCKEYKRGGLRLHNSLIEHGQRNAQCPKCKIHIOREGCDDHMTCSQCNTPNFCRGGERKKYRFGDHTNSLIFGCKRYRLPERPHLRILVSGVACGLFIAPLIMVGLALCAIVAIYEELIKTYNLISGRTRMOTQILAQVPLSDMYCLKQVETICISVLLPL"
join(18776..18917,20080..20134,21625..21732,31415..31579,50130..50331,54909..54980,56337..57191)
/locus_tag="RPl-84N20.1-001"
join(18776..18917,20080..20134,21625..21732,31415..31579,50130..50331,54909..54980,56337..57191)
/locus_tag="RPl-84N20.1-001"
/note="match: ESTs: AA772709 AA776581 AI280328 AI888256 AA4669120 AM859482 BE891363 BE906781 R22137"
join(21665..21732,31415..31579,50130..50331,54909..54980,56337..56410)
/locus_tag="RPl-84N20.1-003"
/standard_name="OTHUMP00000017130"
/note="match: proteins: Q9BX47"
/codon_start=1
/protein_id="CA123000.1"
/db_xref="GI:56204338"
/db_xref="UniProt/TREMBL:Q5TCA3"
/translation="MAPCTETKIFVLCDBOIQCTPCQFWCKCHSPHMEGVNCKEYKQKDLRLHNSLIEHGQRNAQCPKCKIHIOREGCDDHMTCSQCNTPNFCRGGERKKYRFGDHTNSLIFGCKRYRLPERPHLRILVSGVACGLFIAPLIMVGLALCAIVAVGLFPIFYCCCKKORRSRTGMHM"
57166..57171
/locus_tag="RPl-84N20.1-001"
57171..57176
/locus_tag="RPl-84N20.1-001"
57192
/locus_tag="RPl-84N20.1-001"
57192..127443,AL121938.10:41776..41891,AL121938.10:50816..50964,AL121938.10:69980..70081,AL121938.10:75415..75415,AL121938.10:78796..78856,AL121938.10:84532..85190)
/gene="TPD52L1"
/locus_tag="RPl-16705.1-001"

gene
 join(127206..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:75415..75453,AL121938.10:84532..85190)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-003"
 join(127206..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:75415..75453,AL121938.10:78796..78856,
 AL121938.10:84532..85190)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-001"
 /note="match: ESTs: AA102528 AA948661 A1191454 A1193115
 AA608564 AA633641 AA702528 AA948661 A1191454 A1193115
 A1215598 A1222450 A1278449 A1286037 A14923371
 A1561287 A1807361 A1810226 A1826582 A1826585 A1912472
 A120549 AW118159 AW189554 AW245382 AW389720 BE378294
 BE409865 BR037260 BG328591 BG332781 BG393535 BG490930
 B1086202 B1560760 BM542111 BM809770 BM913750 F66007 F32726
 F6468 N99206 T68330 T89899 T93647 W93489
 match: CDNA: AF004428 AF004429 AF004430 U44427 U44428
 Y07757"
 join(127206..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:75415..75453,AL121938.10:84532..85190)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-003"
 /note="match: ESTs: BG110043.1 B158982.1 B1904596.1
 match: CDNA: BC002375.1"
 join(127315..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:84532..85196)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-004"
 join(127315..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:75415..75453,AL121938.10:84532..85196)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-004"
 /note="match: ESTs: BE407159.1 BE619243.1 BE669833.1
 BE677422.1 BG770463.1 BT091199.1 BM802217.1 BM809770.1
 match: CDNA: AF208012.1 BC002375.1"
 join(127425..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:75415..75453,AL121938.10:84532..85411)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-003"
 /standard_name="OTTHUMP0000017135"
 /codon_start=1
 /protein_id="CA123002.1"
 /db_xref="GI:56204340"
 /db_xref="InterPro:IPR007327"
 /db_xref="UniProt/TrEMBL:Q9BU06"
 /translation="MEAAQGLLETP/LQGTDEDAVASADFSMTLSEEEKELKAEIV
 QLESDITLRQVLSAKERHLEIKQKGMNIMLEKQNFPSKSHMDQTTAYKKTHT
 LSHAGOKATAAFSNVGTATSKFGDMSYIRHSISNPMARK"
 join(127425..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:84532..85411)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-004"
 /standard_name="OTTHUMP0000017136"
 /codon_start=1
 /protein_id="CA123003.1"
 /db_xref="GI:56204341"
 /db_xref="InterPro:IPR007327"
 /db_xref="UniProt/TrEMBL:Q9C054"
 /translation="MEAAQGLLETP/LQGTDEDAVASADFSMTLSEEEKELKAEIV
 QLESDITLRQVLSAKERHLEIKQKGMNIMLEKQNFPSKSHMDQTTAYKKTHT
 LSHAGOKATAAFSNVGTATSKFGDMSYIRHSISNPMARK"
 join(127425..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:84532..85411)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-003"
 /standard_name="OTTHUMP0000017135"
 /codon_start=1
 /protein_id="CA123002.1"
 /db_xref="GI:56204340"
 /db_xref="InterPro:IPR007327"
 /db_xref="UniProt/TrEMBL:Q9BU06"
 /translation="MEAAQGLLETP/LQGTDEDAVASADFSMTLSEEEKELKAEIV
 QLESDITLRQVLSAKERHLEIKQKGMNIMLEKQNFPSKSHMDQTTAYKKTHT
 LSHAGOKATAAFSNVGTATSKFGDMSYIRHSISNPMARK"
 join(127425..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:84532..85411)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-004"
 /standard_name="OTTHUMP0000017136"
 /codon_start=1
 /protein_id="CA123003.1"
 /db_xref="GI:56204341"
 /db_xref="InterPro:IPR007327"
 /db_xref="UniProt/TrEMBL:Q9C054"
 /translation="MEAAQGLLETP/LQGTDEDAVASADFSMTLSEEEKELKAEIV
 QLESDITLRQVLSAKERHLEIKQKGMNIMLEKQNFPSKSHMDQTTAYKKTHT
 LSHAGOKATAAFSNVGTATSKFGDMSYIRHSISNPMARK"
 join(127425..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:84532..85411)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-003"
 /standard_name="OTTHUMP0000017135"
 /codon_start=1
 /protein_id="CA123002.1"
 /db_xref="GI:56204340"
 /db_xref="InterPro:IPR007327"
 /db_xref="UniProt/TrEMBL:Q9BU06"
 /translation="MEAAQGLLETP/LQGTDEDAVASADFSMTLSEEEKELKAEIV
 QLESDITLRQVLSAKERHLEIKQKGMNIMLEKQNFPSKSHMDQTTAYKKTHT
 LSHAGOKATAAFSNVGTATSKFGDMSYIRHSISNPMARK"
 join(127425..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:84532..85411)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-004"
 /standard_name="OTTHUMP0000017136"
 /codon_start=1
 /protein_id="CA123003.1"
 /db_xref="GI:56204341"
 /db_xref="InterPro:IPR007327"
 /db_xref="UniProt/TrEMBL:Q9C054"
 /translation="MEAAQGLLETP/LQGTDEDAVASADFSMTLSEEEKELKAEIV
 QLESDITLRQVLSAKERHLEIKQKGMNIMLEKQNFPSKSHMDQTTAYKKTHT
 LSHAGOKATAAFSNVGTATSKFGDMSYIRHSISNPMARK"
 join(127425..127443,AL121938.10:41776..41891,
 AL121938.10:50816..50964,AL121938.10:65980..70081,
 AL121938.10:84532..85411)
 /gene="TPD52L1"
 /locus_tag="RPI-16705.1-003"
 /standard_name="OTTHUMP0000017135"
 /codon_start=1
 /protein_id="CA123002.1"
 /db_xref="GI:56204340"
 /db_xref="InterPro:IPR007327"
 /db_xref="UniProt/TrEMBL:Q9BU06"
 /translation="MEAAQGLLETP/LQGTDEDAVASADFSMTLSEEEKELKAEIV
 QLESDITLRQVLSAKERHLEIKQKGMNIMLEKQNFPSKSHMDQTTAYKKTHT
 LSHAGOKATAAFSNVGTATSKFGDMSYIRHSISNPM

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGAACAACTCT 25
 Db 111048 AAAAAAAAAATCGAACAACTCT 111072

RESULT 16
 AC019270/c
 LOCUS AC019270 154957 bp DNA linear PRI 14-MAR-2002
 DEFINITION Homo sapiens chromosome 8, clone RP11-10C8, complete sequence.
 ACCESSION AC019270
 VERSION AC019270.12 GI:18767544
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.

REFERENCE 1 (bases 1 to 154957)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 8, clone RP11-10C8
 Unpublished
 2 (bases 1 to 154957)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
 Boguslavsky, L., Bouhgalter, B., Brown, A., Burkett, G., Castele, A.,
 Choapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearellano, K., Dewar, K., Domino, M., Doyle, M., Feneclor, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGarr, A., McKernan, K.,
 McPheders, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Leary, T. M., Peterson, K.,
 Pierre, N., Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A. and Zody, W.

TITLE Direct Submission
 JOURNAL Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 154957)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choapel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazeres, R.,
 Landers, T., Lehoczy, J., Levine, R., Liu, G., Macleane, C.,
 Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strause, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, W.

TITLE Direct Submission
 JOURNAL Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 4 (bases 1 to 154957)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Bouhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choapel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazeres, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Macleane, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strause, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, W.

TITLE Direct Submission
 JOURNAL Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 20, 2002 this sequence version replaced gi:18252020.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 Project Information
 Center project name: U3046
 Center clone name: 10_C8

FEATURES
 source

 Location/Qualifiers
 1. 154957
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-10C8"
 /clone_lib="RPCT-11 Human Male BAC"
 2216..2329
 /rpt_family="MER113"
 complement(2502..2647)
 /rpt_family="MIR3"
 3423..3731
 /rpt_family="Alu"
 repeat_region
 4765..4905
 /rpt_family="MIR"
 repeat_region
 5524..5593
 /rpt_family="(TA)n"
 repeat_region
 6476..6592
 /rpt_family="FLAM_C"
 repeat_region
 6593..6616
 /rpt_family="(CAT)n"
 repeat_region
 6617..6634
 /rpt_family="FLAM_C"
 repeat_region
 complement(7783..7846)
 /rpt_family="MER94"
 7858..8272
 /rpt_family="LTR33"
 complement(9296..9374)
 /rpt_family="MIR"
 repeat_region
 complement(9375..9676)
 /rpt_family="Alu"
 repeat_region
 complement(9677..9870)
 /rpt_family="MIR"
 10733..11113
 /rpt_family="MSTA"

repeat_region	complement(11165..11453)
repeat_region	/rpt_family="AluSg"
repeat_region	11881..12074
repeat_region	/rpt_family="MSTB1"
repeat_region	12076..12202
repeat_region	/rpt_family="FLAM_C"
repeat_region	12216..12456
repeat_region	/rpt_family="MSTB1"
repeat_region	12601..12909
repeat_region	/rpt_family="AluSc"
repeat_region	13865..13971
repeat_region	/rpt_family="MTR3"
repeat_region	13972..13992
repeat_region	/rpt_family="AT_rich"
repeat_region	14119..14422
repeat_region	/rpt_family="MLT1F"
repeat_region	14458..14583
repeat_region	/rpt_family="MLT1F"
repeat_region	complement(15103..15138)
repeat_region	/rpt_family="MTR"
repeat_region	15630..15932
repeat_region	/rpt_family="AluY"
repeat_region	16793..16823
repeat_region	/rpt_family="(CA)n"
repeat_region	17216..17279
repeat_region	/rpt_family="Tigge3(colem)"
repeat_region	17376..17398
repeat_region	/rpt_family="AT_rich"
repeat_region	17399..17507
repeat_region	/rpt_family="Tigge3(colem)"
repeat_region	complement(18015..18184)
repeat_region	/rpt_family="MER5A"
repeat_region	complement(18283..18615)
repeat_region	/rpt_family="AluY"
repeat_region	18784..18809
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(18850..18960)
repeat_region	/rpt_family="FLAM_C"
repeat_region	complement(19002..20375)
repeat_region	/rpt_family="L2"
repeat_region	complement(21297..21594)
repeat_region	/rpt_family="AluJo"
repeat_region	complement(23766..24061)
repeat_region	/rpt_family="AluSg"
repeat_region	24391..24412
repeat_region	/rpt_family="AT_rich"
repeat_region	complement(24568..24737)
repeat_region	/rpt_family="AluJo"
repeat_region	24738..24767
repeat_region	/rpt_family="(TTTG)n"
repeat_region	complement(24768..24891)
repeat_region	/rpt_family="AluJo"
repeat_region	complement(25207..26166)
repeat_region	/rpt_family="L2"
repeat_region	complement(26170..26373)
repeat_region	/rpt_family="MTR"
repeat_region	complement(26384..26635)
repeat_region	/rpt_family="L2"
repeat_region	27277..27325
repeat_region	/rpt_family="AT_rich"
repeat_region	27949..28084
repeat_region	/rpt_family="MTR"
repeat_region	29469..29606
Query Match	93.6%; Score 23.4; DB 8; Length 154957;
Best Match	96.0%; Pred. No. 1.9e+02;
Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT	17
AC147311	
LOCUS	
DEFINITION	161659 bp DNA linear HTG 07-NOV-2003
ACCESSION	AC147311
VERSION	Pan troglodytes chromosome UNK clone RH43-51E21, *** SEQUENCING IN PROGRESS ***, 26 unordered pieces.
KEYWORDS	AC147311 AC147311.1 GI:38198851
SOURCE	HTG; HTGS PHASE1.
ORGANISM	Pan troglodytes (chimpanzee)
	Pan troglodytes
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarcheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
REFERENCE	1 (bases 1 to 161659)
AUTHORS	Wilson,R.K.
TITLE	The sequence of Pan troglodytes clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 161659)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (07-NOV-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: C_PR051E21
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1193: contig of 1193 bp in length
*
*      1194      1293: gap of unknown length
*      1294      2365: contig of 1072 bp in length
*      2366      2465: gap of unknown length
*      2466      4408: contig of 1943 bp in length
*      4409      4508: gap of unknown length
*      4509      5733: contig of 1225 bp in length
*      5734      5833: gap of unknown length
*      5834      7359: contig of 1526 bp in length
*      7360      7459: gap of unknown length
*      7460      10297: contig of 2838 bp in length
*      10298      10397: gap of unknown length
*      10398      13797: contig of 3400 bp in length
*      13798      13897: gap of unknown length
*      13898      16242: contig of 2345 bp in length
*      16243      16342: gap of unknown length
*      16343      19925: contig of 3583 bp in length
*      19926      20025: gap of unknown length
*      20026      24198: contig of 4173 bp in length
*      24199      24298: gap of unknown length
*      24299      28451: contig of 4153 bp in length
*      28452      28551: gap of unknown length
*      28552      33232: contig of 4681 bp in length
*      33233      33332: gap of unknown length
*      33333      36785: contig of 3453 bp in length
*      36786      36885: gap of unknown length
*      36886      43285: contig of 6400 bp in length
*      43286      43385: gap of unknown length
*      43386      48870: contig of 5485 bp in length
*      48871      48970: gap of unknown length
*      48971      55808: contig of 6838 bp in length
*      55809      55908: gap of unknown length
*      55909      62392: contig of 6484 bp in length
*      62393      62492: gap of unknown length

```


ORGANISM	COMMENT
Homo sapiens	On Jan 25, 2003 this sequence version replaced gi:27819428.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;	gc-help@bcm.tmc.edu
Homindae; Homo.	
1 (bases 1 to 161841)	CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.
Munzy D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C., Albrooks S.L., Amaralunge H.C., Are J.R., Ayele M., Banks T., Barbata U., Benton J., Bimage K., Blankenburg K., Bonini D., Bouck J., Bowie S., Brileva M., Brown E., Brown M., Bryant N.P., Butay C., Burch P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R., Chen Z., Chiu D., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R., Davila L.M., Davis C., Day-Carroll L., Dederich D.A., Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J., Earhart C., Edgar D., Edwards C.C., Elhaj C., Emelling S., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Francis P., Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K., Han J., Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J., Hernandez O., Hodgson A., Hognes M., Holloway C., Hollins B., Homsi F., Howard S., Huber J., Hulik S., Hume J., Ioshikhes I., Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S., Joudan S., Karlsson E., Kelly S., Khan U., King L., Korvah J., Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lee E., Lewis L.C., Lewis L., Li J., Li Z., Lichtarge O., Lien C., Liu J., Liu W., Louised H., Lozado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J., Maheshwari M., Mapa P., Marondel I., Martin R., Martindale A., Martinez E., Massey E., Mawinney E., McLeod M.P., Meador M., Mel G., Mescher S., Metker M., Miller A., Miner G., Miner Z., Mitchell T., Mohabbat K., Montgomery K.T., Morgan M., Morris S., Moser M., Neal D., Nelson D., Newton J., Newton N., Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokenko S., Ogih M., Okunolu G., Orgunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y., Rives M., Rojas A., Rojudoan I., Rolfe M., Ruiz S., Saverly G., Scherer S., Scott G., Shen H., Shim C., Shoshchani N., Sisson I., Sodergren E., Sonalike T., Sparks A., Stanley H., Stone H., Sutton A., Svatek A., Taylor P., Tamerisa A., Tamerisa K., Tang H., Taney J., Taylor C., Taylor P., Telford B., Thomas N., Thomas S., Umami K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q., Wang S., Ward-Moore S., Warren R., Washington C., Watlington S., Williams G., Williamson A., Wleczek R., Wooden S., Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Kucherlapati R., Weinstein G. and Gibbs R.	
Direct Submission	
Unpublished	
2 (bases 1 to 161841)	repeat_region
Worley K.C.	repeat_region
Direct Submission	repeat_region
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	repeat_region
3 (bases 1 to 161841)	repeat_region
Worley K.C.	repeat_region
Direct Submission	repeat_region
Submitted (22-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	repeat_region
4 (bases 1 to 161841)	repeat_region
Worley K.C.	repeat_region
Direct Submission	repeat_region
Submitted (25-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	repeat_region
5 (bases 1 to 161841)	repeat_region
Worley K.C.	repeat_region
Direct Submission	repeat_region
Submitted (29-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	repeat_region

FEATURES

source

LOCATION/Qualifiers

1..161841

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="12"

/clone="RP11-114F3"

/complement(1..2000)

/note="overlaps bases 1..2000 of clone AC136977"

/function="clone overlap"

838..862

/rpt_family="AT_rich"

/complement(936..1150)

/rpt_family="AluYb"

/complement(1639..1650)

/rpt_family="AluX"

1651..1694

/rpt_family="(TG)n"

/complement(1695..1975)

/rpt_family="AluX"

2005..2217

/rpt_family="MIR"

/complement(2396..2616)

/rpt_family="L1MA2"

/complement(2629..2798)

/rpt_family="L1MD2"

/complement(2815..3141)

/rpt_family="MER7A"

3956..3993

/rpt_family="AT_rich"

/complement(4063..4220)

/rpt_family="FAM"

4283..4311

/rpt_family="AT_rich"

4780..4802

repeat_region

```

repeat_region      /rpt_family="AT_rich"
                    6237..6261
repeat_region      /rpt_family="AT_rich"
                    7248..7585
repeat_region      /rpt_family="AluSg"
                    7867..9243
repeat_region      /rpt_family="L1PA16"
                    9244..9552
repeat_region      /rpt_family="AluY"
                    9553..10132
repeat_region      /rpt_family="L1PA16"
                    10226..10428
repeat_region      /rpt_family="L1PA16"
                    10554..10864
repeat_region      /rpt_family="AluUo"
                    11316..11760
repeat_region      /rpt_family="L2"
                    11761..12056
repeat_region      /rpt_family="AluSg"
                    12057..12403
repeat_region      /rpt_family="L2"
                    12382..12674
repeat_region      /rpt_family="L2"
                    13032..13032
repeat_region      /rpt_family="MER30"
                    14051..14333
repeat_region      /rpt_family="AluX"
                    14355..14539
repeat_region      /rpt_family="AluSg/x"
                    15160..15531
repeat_region      /rpt_family="MER20B"
                    15540..15698
repeat_region      /rpt_family="L2"
                    16385..16424
repeat_region      /rpt_family="L2"
                    16969..17173
repeat_region      /rpt_family="L2"
                    17193..17485
repeat_region      /rpt_family="AluSg"
                    17513..17787
repeat_region      /rpt_family="AluUo"
                    18942..18964
repeat_region      /rpt_family="AT_rich"
                    19765..20435
repeat_region      /rpt_family="L1MB7"
                    20436..20738
repeat_region      /rpt_family="AluUo"

Query Match      93.6%; Score 23.4; DB 8; length 161841;
Best Local Similarity 96.0%; Pred.No.1.9e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 91936 AAAAAAAAAATCGCAACAAATCT 91960

RESULT 19
AC021786 163554 bp DNA linear HTG 30-MAR-2000
LOCUS Homo sapiens clone RP11-21H8, WORKING DRAFT SEQUENCE, 8 unordered
pieces.
AC021786
AC021786.2 GI:7341964
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 163554)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-21H8

```

```

JOURNAL
REFERENCE
AUTHORS
unpublished
2 (bases 1 to 163554)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavsky,I., Bouckhalter,B., Brown,A., Burtett,G., Castle,A.,
Choapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardina,S., Grant,G., Hago,B., Healdord,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kam,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,C., Liu,G., Locke,R.,
Macdonald,P., Marquis,N., McGwan,P., McGuirk,A., McKernan,K.,
Mcneale,R., Meldrum,J., Menue,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wymann,D., Ye,W.J.,
Zimmer,A. and Zody,M.

Direct Submission
Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 30, 2000 this sequence version replaced gi:6721341.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information
Center project name: L4019
Center clone name: 21_H_8
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158445 bases at least Q40
Consensus quality: 161065 bases at least Q30
Insert size: 167000; agarose-fp
Insert size: 162854; sum-of-contigs
Quality coverage: 5.3 in Q20 bases; agarose-fp
Quality coverage: 5.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1174: contig of 1174 bp in length
* 1175 1274: gap of 100 bp
* 1275 4788: contig of 3514 bp in length
* 4788 4888: gap of 100 bp
* 4888 14873: contig of 9985 bp in length
* 14873 14973: gap of 100 bp
* 14974 32556: contig of 17583 bp in length
* 32557 32557: gap of 100 bp
* 32557 32576: contig of 19920 bp in length
* 32576 52676: gap of 100 bp
* 52677 73204: contig of 20528 bp in length
* 73205 73304: gap of 100 bp
* 73305 109834: contig of 36530 bp in length
* 109835 109934: gap of 100 bp
* 109935 163554: contig of 53620 bp in length.
Location/Qualifiers
1..163554
/organism="Homo sapiens"
/mol_type="genomic DNA"
FEATURES
source

```

```

/db xref="taxon:3606"
/clone="RP11-21H6"
/clone_lib="RPCT-11 Human Male BAC"
1..1174
misc_feature
/note="assembly_fragment"
1175..1274
/estimated_length=100
1275..4786
/note="assembly_fragment"
4789..4888
/estimated_length=100
4889..14873
/note="assembly_fragment"
14874..14973
/estimated_length=100
14974..32556
/note="assembly_fragment"
32557..32656
/estimated_length=100
32657..52576
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
52577..52676
/estimated_length=100
52677..73205
/note="assembly_fragment"
73205..73304
/estimated_length=100
73305..109834
/note="assembly_fragment"
109835..109934
/estimated_length=100
109935..163554
/note="assembly_fragment"
clone_end:r7
vector_side:right"

ORIGIN
Query Match 93.6%; Score 23.4; DB 14; Length 163554;
Best Local Similarity 96.0%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
    |||||
Db 9454 AAAAAAAAAATCGCAAAAATCT 9478

RESULT 20
AC008949/c AC008949 163721 bp DNA linear PRI 23-AUG-2001
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2335024, complete sequence.
ACCESSION AC008949
VERSION AC008949.9 GI:15281188
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 163721)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 163731)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163731)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell

```

```

COMMENT
On Aug 23, 2001 this sequence version replaced gi:13699462.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-hgsc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence:
Estimated Total Number of Errors is 0.5.

FEATURES
    source
        1..163731
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="CTD-2335024"

ORIGIN
Query Match      93.6%; Score 23.4; DB 8; Length 163731;
Best Local Similarity 96.0%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGCAACCAATCT 25
        |||||||||||||||||||
        AAAAAAAAAATCGCAACCAATCT 61397

Db      61421 AAAAAAAAAATCGCAACCAATCT 61397

RESULT 21
AL591663/c      166517 bp      DNA      linear      HTG 19-DEC-2001
LOCUS
DEFINITION
Homo sapiens chromosome 10 clone RP13-206M19, WORKING DRAFT
SEQUENCE, 28 unordered pieces.
AL591663
AL591663.2 GI:14529990
HTG; HTGS PHASE1; HTGS _DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1
Burton,J.
Direct Submission
Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 22, 2001 this sequence version replaced gi:14268334.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: hb206M19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 150526 bases at least Q40
Consensus quality: 156647 bases at least Q30
Consensus quality: 160260 bases at least Q20
Insert size: 163817; sum-of-contigs
Insert size: 256494; 24.1% error; agarose-fp
Quality coverage: 3.02x in Q20 bases; sum-of-contigs Quality
coverage: 2.23x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```
* 1 6863: contig of 6863 bp in length
* 6864 6963: gap of 100 bp
* 6964 14905: contig of 7942 bp in length
* 14906 15005: gap of 100 bp
* 15006 17487: contig of 2482 bp in length
* 17488 17588: gap of 100 bp
* 17588 23573: contig of 5986 bp in length
* 23574 23674: gap of 100 bp
* 23674 32984: contig of 9311 bp in length
* 32985 33084: gap of 100 bp
* 33085 37288: contig of 4203 bp in length
* 37288 43136: contig of 5749 bp in length
* 43137 43236: gap of 100 bp
* 43237 47819: contig of 4583 bp in length
* 47820 47919: gap of 100 bp
* 47920 50691: contig of 2772 bp in length
* 50692 50791: gap of 100 bp
* 50792 57010: contig of 6219 bp in length
* 57011 57111: gap of 100 bp
* 57111 60822: contig of 3712 bp in length
* 60823 60923: gap of 100 bp
* 60923 63855: contig of 2933 bp in length
* 63856 63956: gap of 100 bp
* 63956 74630: contig of 10675 bp in length
* 74631 74730: gap of 100 bp
* 74731 83250: contig of 8520 bp in length
* 83251 83350: gap of 100 bp
* 83351 93216: contig of 9866 bp in length
* 93217 93316: gap of 100 bp
* 93317 98003: contig of 4687 bp in length
* 98004 98103: gap of 100 bp
* 98104 107042: contig of 8939 bp in length
* 107043 107142: gap of 100 bp
* 107143 111209: contig of 4067 bp in length
* 111210 111309: gap of 100 bp
* 111310 117200: contig of 5891 bp in length
* 117201 117300: gap of 100 bp
* 117301 121070: contig of 3770 bp in length
* 121071 121170: gap of 100 bp
* 121171 128898: contig of 7728 bp in length
* 128899 128998: gap of 100 bp
* 128999 131416: contig of 2418 bp in length
* 131417 131516: gap of 100 bp
* 131517 137669: contig of 6153 bp in length
* 137670 137769: gap of 100 bp
* 137770 145371: contig of 7602 bp in length
* 145372 145471: gap of 100 bp
* 145472 148522: contig of 3051 bp in length
* 148523 148622: gap of 100 bp
* 148623 151178: contig of 2556 bp in length
* 151179 151278: gap of 100 bp
* 151279 162775: contig of 11497 bp in length
* 162776 162875: gap of 100 bp
* 162876 16517: contig of 3642 bp in length.
```

FEATURES

source

```
1..16517
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone_id="RPC1-13.1"
1..6863
/note="assembly_fragment:00213
fragment_chain:1"
misc_feature
6964..14905
/note="assembly_fragment:01337
fragment_chain:1"
misc_feature
15006..17487
/note="assembly_fragment:00737
fragment_chain:1"
misc_feature
17588..23573
```

```
/note="assembly_fragment:01703
fragment_chain:1"
misc_feature
23674..32984
/note="assembly_fragment:01082
fragment_chain:1"
misc_feature
33085..37288
/note="assembly_fragment:01089
fragment_chain:1"
misc_feature
37388..43136
/note="assembly_fragment:01752
fragment_chain:1"
misc_feature
43237..47819
/note="assembly_fragment:00709
fragment_chain:1"
misc_feature
47920..50691
/note="assembly_fragment:01258
fragment_chain:1"
misc_feature
50792..57010
/note="assembly_fragment:01159
fragment_chain:2"
misc_feature
57111..60822
/note="assembly_fragment:01165
fragment_chain:2"
misc_feature
60923..63855
/note="assembly_fragment:01793
fragment_chain:2"
misc_feature
63956..74630
/note="assembly_fragment:01144
fragment_chain:2"
misc_feature
74731..83250
/note="assembly_fragment:00135
fragment_chain:3"
misc_feature
83351..93216
/note="assembly_fragment:01427
fragment_chain:3"
misc_feature
93317..98003
/note="assembly_fragment:00320
fragment_chain:3"
misc_feature
98104..107042
/note="assembly_fragment:01115
fragment_chain:4"
misc_feature
107143..117200
/note="assembly_fragment:00036
fragment_chain:4"
misc_feature
117301..117300
/note="assembly_fragment:01713
fragment_chain:4"
misc_feature
117301..121070
/note="assembly_fragment:00152"
121171..128898
/note="assembly_fragment:00190"
128999..131416
/note="assembly_fragment:00362"
131517..137669
/note="assembly_fragment:00390"
137770..145371
/note="assembly_fragment:00669"
145472..148522
/note="assembly_fragment:00694"
148623..151178
/note="assembly_fragment:00793"
151279..162775
/note="assembly_fragment:00836"
162876..16517
/note="assembly_fragment:01551"
```

ORIGIN

```
Query Match 93.6%; Score 23.4; DB 14; Length 16517;
Best Local Similarity 96.0%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACAAATCT 25
|||||
```

DB 138850 AAAAAAAAAATCACAACAACTCT 138826

RESULT 22
AC021706/c

LOCUS
DEFINITION AC021706 168833 bp DNA linear HTG 08-APR-2000
Homosapiens clone RP11-388017, WORKING DRAFT SEQUENCE, 33
unorderd pieces.
AC021706

ACCESSION
AC021706.3 GI:7528116

VERSION
HTG; HTGS PHASE1; HTGS_DRAFT.

KEYWORDS
SOURCE
ORGANISM
Homosapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 Birren, B., Linton, L., Nusbaum, C., and Lander, E.
2 (bases 1 to 168833)
Unpublished
2 (bases 1 to 168833)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Bouhgalter, B., Brown, A., Burkett, G., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Deaellano, K., Dewar, K., Domino, M., Doyle, M., Fenesfor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Garlyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGuirk, A., McKernan, K.,
McPheters, R., Meldrum, J., Meneses, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vasiliou, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W., J.,
Zimmer, A., and Zody, M.

Direct Submission
Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced gi:6899688.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L5882
Center clone name: 388 O 17

Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149025 bases at least Q40
Consensus quality: 158058 bases at least Q30
Consensus quality: 162087 bases at least Q20
Insert size: 18600; agarose-fp
Insert size: 165633; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1124: contig of 1124 bp in length
1125 1224: gap of 100 bp
1225 2246: contig of 1022 bp in length
1227 2346: gap of 100 bp
2347 3895: contig of 1549 bp in length
3896 3996: gap of 100 bp
3996 5068: contig of 1073 bp in length
5069 5168: gap of 100 bp
5169 6557: contig of 1389 bp in length
6558 6658: gap of 100 bp
6658 8302: contig of 1644 bp in length
8302 8402: gap of 100 bp
8402 9758: contig of 1357 bp in length
9759 9859: gap of 100 bp
9859 12176: contig of 2318 bp in length
12177 12276: gap of 100 bp
12277 14646: contig of 2370 bp in length
14647 14746: gap of 100 bp
14747 17380: contig of 2634 bp in length
17381 17480: gap of 100 bp
17481 19793: contig of 2313 bp in length
19794 19893: gap of 100 bp
19894 22375: contig of 2482 bp in length
22376 22475: gap of 100 bp
22476 24401: contig of 1926 bp in length
24402 24501: gap of 100 bp
24502 27437: contig of 2936 bp in length
27438 27537: gap of 100 bp
27538 29420: contig of 1883 bp in length
29421 3066: gap of 100 bp
3066 33066: contig of 3356 bp in length
33067 33166: gap of 100 bp
33167 37225: contig of 4059 bp in length
37226 37325: gap of 100 bp
37326 40161: contig of 2836 bp in length
40162 40261: gap of 100 bp
40262 44797: contig of 4536 bp in length
44798 44897: gap of 100 bp
44898 51246: contig of 6349 bp in length
51247 51346: gap of 100 bp
51347 58352: contig of 6906 bp in length
58353 58352: gap of 100 bp
58353 64134: contig of 5782 bp in length
64135 64235: gap of 100 bp
64235 69861: contig of 5627 bp in length
69862 73026: gap of 100 bp
73027 73126: gap of 100 bp
73127 80088: contig of 6962 bp in length
80089 80188: gap of 100 bp
80189 87519: contig of 7331 bp in length
87520 87619: gap of 100 bp
87620 93596: contig of 5977 bp in length
93597 93696: gap of 100 bp
93697 103354: contig of 9558 bp in length
103355 103354: gap of 100 bp
103355 115235: contig of 11881 bp in length
115236 115335: gap of 100 bp
115336 125315: contig of 9880 bp in length
125316 125415: gap of 100 bp
125416 138402: contig of 12987 bp in length
138403 138502: gap of 100 bp
138503 151752: contig of 13250 bp in length
151753 151852: gap of 100 bp
151853 168833: contig of 16981 bp in length.

FEATURES

source

Location/Qualifiers
1..168833
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-388017"
/clone_id="RPCT-11 Human Male BAC"
1..1124
misc_feature

```

gap                                     /note="assembly_fragment"
                                     1125..1224
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     1225..2246
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     2247..2346
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     2347..3895
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     3896..3995
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     3996..5068
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     5069..5168
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     5169..6557
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     6558..6657
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     6658..8301
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     8302..8401
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     8402..9758
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     9759..9858
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     9859..12176
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     12177..12276
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     12277..14646
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     14647..14746
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     14747..17380
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     17381..17480
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     17481..19793
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     19794..19893
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     19894..22375
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     22376..22475
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     22476..24401
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     24402..24501
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     24502..27437
                                     /estimated_length=100
gap                                     /note="assembly_fragment"
                                     27438..27537
                                     /estimated_length=100
misc_feature                           /note="assembly_fragment"
                                     27538..29420
                                     /estimated_length=100
Query Match                          93.6%: Score 23.4; DB 14; Length 168833;
Best Local Similarity 96.0%: Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OR 1 AAAAAAAAAATCGCAACCAATCT 25
    |||||
Db 6904 AAAAAAAAAATCGCAACCAATCT 6880
    |||||
RESULT 23
AL353616 AL353616 169993 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-274B18 on chromosome 9 Contains

```

```

ACCESSION AL353616
VERSION HTG; MGC34760; PGM5; PGMRP; SART2.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.
1 (bases 1 to 169993)
Leongamornlert, P.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jan 19, 2002 this sequence version replaced GI:12584384.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-274B18 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: DBACe3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
FEATURES
Source
1..169993
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-274B18"
/clone_id="RPCI-11.1"
1
/note="Clone left end: RP11-274B18"
join(AL353608.20:136408..171921,AL161457.13:73550..107818,
15636..17130)
/gene="PGM5"
/locus_tag="RP11-274B18.5-001"
join(AL353608.20:157647..157809,AL353608.20:163846..163992,
AL353608.20:166911..167036,AL353608.20:170982..171172,
AL353608.20:171767..171921,AL161457.13:73550..73665,
AL161457.13:87875..88010,AL161457.13:92322..92505,
AL161457.13:107684..107818,15636..17130)
/gene="PGM5"
/locus_tag="RP11-274B18.5-001"
/product="phosphoglucosaminidase 5"
/note="match: ESTs: AA631141.1 AA706788.1 AL046941.1
AV396258.1 AV723073.1 AW652952.1 AW652958.1 BB264129.2
BB641396.1 BB628028.1 BB641261.1 BE221369.1 BE235707.1
BF058433.1 BF060711.1 BF230880.1 BF306296.1 BF306987.1
BG254110.1 B1346230.1 B1847349.1 BM363295.1 BM363433.1
BM364512.1 BM366644.1 BM699018.1 BM975035.1 BQ001679.1

```


JOURNAL
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 171732)

REFERENCE
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gattorna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Major, J., Margrie, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Meldrum, J., Menush, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, W., Travis, N., Triggillo, J., Vaasiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
 Direct Submission

JOURNAL
 Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Nov 22, 2001 this sequence version replaced g1:16118193.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 16161
 Center clone name: 250_P_18
 ----- Location/Qualifiers
 1. 171732
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-250P18"
 /clone_lib="RPC1-11 Human Male BAC"
 complement(306..380)
 /rpt_family="MSTC"
 repeat_region
 404..433
 /rpt_family="AT_rich"
 complement(1195..1281)
 /rpt_family="Zaphod"
 3566..3597
 /rpt_family="AT_rich"
 complement(3799..3904)
 /rpt_family="L2"
 repeat_region
 4689..4729
 /rpt_family="L2"
 repeat_region
 4846..4914
 /rpt_family="(TG)n"
 complement(4916..5427)
 /rpt_family="L1P82"
 5429..5520
 /rpt_family="L1P82"
 repeat_region
 5728..5766
 /rpt_family="L1P82"
 repeat_region
 5832..5982
 /rpt_family="AT_rich"
 repeat_region
 5986..7066
 /rpt_family="L1ME1"
 repeat_region

/rpt_family="L1ME1"
 7067..7361
 /rpt_family="Aluv"
 repeat_region
 7362..7821
 /rpt_family="L1ME1"
 repeat_region
 7811..7876
 /rpt_family="L1ME1"
 repeat_region
 9138..9174
 /rpt_family="(CA)n"
 repeat_region
 9419..9700
 /rpt_family="Aluv"
 repeat_region
 9701..9741
 /rpt_family="(CATR)n"
 repeat_region
 10129..10280
 /rpt_family="MIR"
 repeat_region
 10862..10958
 /rpt_family="MIR"
 repeat_region
 10962..11266
 /rpt_family="Aluv"
 repeat_region
 12275..12434
 /rpt_family="MIR"
 repeat_region
 complement(12795..12886)
 /rpt_family="5S"
 repeat_region
 13221..13245
 /rpt_family="(T)n"
 repeat_region
 13602..13907
 /rpt_family="L2"
 repeat_region
 complement(14878..15223)
 /rpt_family="THEIC"
 repeat_region
 complement(15449..15541)
 /rpt_family="MIR3"
 repeat_region
 15940..15969
 /rpt_family="AT_rich"
 repeat_region
 16167..16309
 /rpt_family="FRAM"
 repeat_region
 16317..16347
 /rpt_family="(GA)n"
 repeat_region
 16379..16481
 /rpt_family="MIR"
 repeat_region
 complement(16625..16791)
 /rpt_family="AluCb"
 repeat_region
 complement(17312..17609)
 /rpt_family="AluSx"
 repeat_region
 complement(17660..18024)
 /rpt_family="MUTIA2"
 repeat_region
 19646..19689
 /rpt_family="L2"
 repeat_region
 20540..20590
 /rpt_family="AT_rich"
 repeat_region
 21152..21204
 /rpt_family="MIR"
 repeat_region
 21222..21367
 /rpt_family="L1ME1"
 repeat_region
 21419..21654
 /rpt_family="L1ME1"
 repeat_region
 complement(21658..21736)
 /rpt_family="MER20"
 repeat_region
 21808..21983
 /rpt_family="L1ME1"
 repeat_region
 complement(21984..22337)
 /rpt_family="L1P45"
 repeat_region
 22338..24087
 /rpt_family="L1ME1"
 repeat_region
 24488..24806
 /rpt_family="AluSx"
 repeat_region
 24904..25311
 /rpt_family="L2"
 repeat_region
 25295..26234
 /rpt_family="L2"
 repeat_region
 complement(26235..26570)
 /rpt_family="AluSg"
 repeat_region
 26571..26631
 /rpt_family="L2"
 repeat_region

```
repeat_region complement(27293..27454)
repeat_region /rpt_family="L2"
repeat_region complement(27494..27913)
repeat_region /rpt_family="L1PA9"
repeat_region complement(28281..29147)
repeat_region /rpt_family="L1PA7"
repeat_region 29486..29768
repeat_region /rpt_family="AluY"
repeat_region 29769..29917
repeat_region /rpt_family="L1PA5"
repeat_region 29919..29959
repeat_region /rpt_family="AT_rich"
repeat_region 30507..30819
repeat_region /rpt_family="AluXs"
repeat_region 31981..32158
repeat_region /rpt_family="MIR"
repeat_region 33793..33818
repeat_region /rpt_family="(TTTC)n"
repeat_region complement(35648..35843)
repeat_region /rpt_family="MIR3"
repeat_region 35924..35948

Query Match 93.6%; Score 23.4; DB 8; Length 17172;
Best Local Similarity 96.0%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 135577 AAAAAAAAAATCGCAACAAATCT 135601

RESULT 25
AL451107 172779 bp DNA linear PRI 18-MAY-2005
LOCUS Human DNA sequence from clone RP11-14C22 on chromosome 10 Contains
DEFINITION the gene for a novel protein similar to lysozyme C-1 (l
4-beta-N-acetylmuramidase C EC 3.2.1.17) (MGC33408), three novel
genes and a supervillin pseudogene (SVIL), complete sequence.
ACCESSION AL451107.6 GI:14148855
VERSION HTG; lysozyme; MGC33408; supervillin; SVIL.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 172779)
Dunn, M.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On May 18, 2001 this sequence version replaced gi:13992064.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-14C22 is from the library RPCI-11.1 constructed by the group
of Pletier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
Draft Sequence Produced by Whitehead Institute/MIT Center for
```

```
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
Location/Qualifiers
1..172779
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RPCI-11.1"
/clone_1ib="RPCI-11.1"
1
/note="Clone left end: RP11-14C22"
complement(join(9406..9483,10605..10718,13161..13238,
15097..15207,16005..16044))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-001"
complement(join(9406..9483,10605..10718,13161..13238,
15097..15207,16005..16044))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-001"
complement(join(9406..9483,10605..10718,13161..13238,
15097..15207,16005..16044))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-001"
/note="match: ESTs: A1671318.1"
complement(14962)
/gene="LYZL2"
/locus_tag="RP11-14C22.2-001"
complement(14969)
/gene="LYZL2"
/locus_tag="RP11-14C22.2-001"
complement(join(14972..15207,16005..16083,29288..29446,
29960..30123,32776..32945))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-002"
complement(join(14972..15207,16005..16083,29288..29446,
29960..30123,32776..32945))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-002"
complement(join(14972..15207,16005..16083,29288..29446,
29960..30123,32776..32945))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-002"
complement(join(14972..15207,16005..16083,29288..29446,
29960..30123,32776..32945))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-001"
complement(join(15138..15207,16005..16083,29288..29446,
29960..30123,32776..32888))
/gene="LYZL2"
/locus_tag="RP11-14C22.2-002"
/standard_name="ORTHUMP0000019400"
/note="match: proteins: P00707 Q8WW16 Q96LF2 Q9CPX3 Q9H1R8
Q9H1R9"
/codon_start=1
/product="lysozyme-like 2"
/protein_id="CAH70607.1"
/db_xref="GI:55664937"
/db_xref="GOA:O6N269"
/db_xref="InterPro:IPR000974"
/db_xref="InterPro:IPR001916"
/db_xref="UniProt/TREMBL:Q6N269"
/translation="MODAPLSCLSPTKKSSVSSADSTKSSASAGTRNLPPQPCUROA
LRMKAGILTLIGCLVTGAESEKIYTRCLAKI FSRAGLDNYGFSIGHWI CMAAYESG
YNTTQYTGIDGSDIDYGI FOINSFAMCGRGKLEKNHGHVACASALVTDDLTDAIICAK
KIVETQGMNTWQGWKRCBGRDLSDMKKDEVS"
```

gene	join	transl	score	db	length	htg
gene	join(32731..32759,33925..33246)					
misc_feature	/locus_tag="RP11-14C22.5-001"					
gene	join(132731..32759,33925..33246)					
misc_feature	/locus_tag="RP11-14C22.5-001"					
product	"putative novel transcript"					
note	"match: ESTs: AW085317.1"					
33224..33229						
locus_tag	"RP11-14C22.5-001"					
33246						
locus_tag	"RP11-14C22.5-001"					
33587						
note	"Clone_right_end: RP11-253D19"					
join	(58672..58894,59752..59808,60513..60718)					
locus_tag	"RP11-14C22.3-001"					
join	(58672..58894,59752..59808,60513..60718)					
locus_tag	"RP11-14C22.3-001"					
product	"putative novel transcript"					
note	"match: ESTs: A1142493.1"					
join	(74749..74839,88445..88592,92422..92511,92598..92684)					
94257..94434,96679..96851,99105..99260,100559..100695,						
101300..101424,101553..101698,103270..103445,						
106705..106849,107596..107872,112481..112636,						
114545..114685,115741..115866,120110..120265,						
120366..120449)						
locus_tag	"RP11-14C22.4-001"					
/pseudo						
join	(174749..74839,88445..88592,92422..92511,92598..92684)					
94257..94434,96679..96851,99105..99260,100559..100695,						
101300..101424,101553..101698,103270..103445,						
106705..106849,107596..107872,112481..112636,						
114545..114685,115741..115866,120110..120265,						
120366..120449)						
locus_tag	"RP11-14C22.4-001"					
note	"match: proteins: O46385 O60611 O60612 O95425 O8K412 O8K413 Q9H1R7"					
/pseudo						
/codon_start=1						
/product	"superf111in (SVIT) pseudogene"					
join	(126434..126519,126602..126923)					
locus_tag	"RP11-14C22.6-001"					
join	(126434..126519,126602..126923)					
locus_tag	"RP11-14C22.6-001"					
/product	"putative novel transcript"					
note	"match: ESTs: A1674082.1"					
126898..126903						
locus_tag	"RP11-14C22.6-001"					
126923						
locus_tag	"RP11-14C22.6-001"					
155081						
note	"Clone_left_end: RP11-330011"					
172779						
note	"Clone_right_end: RP11-14C22"					
ORIGIN						
Query Match	93.6%	Score 23.4	DB 8	Length 172779		
Best Local Similarity	96.0%	Pred. No. 1.9e+02				
Matches 24	Conservative 0	Mismatches 1	Indels 0	Gaps 0		
QY	1	AAAAAAAAAATCGCAACCAATCT	25			
Db	13307	AAAAAAAAAATCGCAACCAATCT	13331			
RESULT 26						
LOCUS AC073894	173491 bp	DNA	linear	HTG 08-JAN-2003		
DEFINITION Homo sapiens chromosome 3 clone RP11-241019, WORKING DRAFT						
ACCESSION AC073894						
VERSION AC073894.15	GI:20335601					
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.						

REFERENCE
AUTHORS
Makayova; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 173491)
Muzny,D.M., Adams,C., Adio-Odola,B., Alt-osman,F.R., Allen,C.,
Alshrooks,S.L., Amarungue,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,J., Benton,J., Blumage,K., Blankenhorn,K., Bonini,D.,
Buck,J., Bowle,S., Briviera,M., Brown,E., Brown,M., Bryant,N.P.,
Bhatia,K., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Deatney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Dreper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N.C., Ford,J., Foster,P., Franz,P.,
Gabisil,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsa,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,U., Liu,M., Louieged,H.,
Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,U.,
Masseyhewari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkewo,S., Ogun,M., Okwoudu,G.,
Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojasa,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtlati,N., Slasson,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Taneay,J., Taylor,C., Taylor,T., Teitford,B., Thomas,N., Thomas,S.,
Uemami,K., Vaequez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 173491)
Worley K.C.
Direct Submission
Submitted (01-JUL-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 173491)
Worley K.C.
Direct Submission
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:16449546.
COMMENT
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: HAZB
Center clone name: RP11-241019
Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-Primer BigDye 4.1% of reads
Chemistry: Dye-terminator Big Dye 5.9% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 164781 bases at least Q40

Consensus quality: 168254 bases at least Q20
 Consensus quality: 169842 bases at least Q20
 Estimated insert size: 173762; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/doc/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1       3719: contig of 3719 bp in length
*       3720       3819: gap of unknown length
*       3820       7606: contig of 3787 bp in length
*       7607       7706: gap of unknown length
*       7707       10227: contig of 2521 bp in length
*       10228      10327: gap of unknown length
*       10328      17034: contig of 6707 bp in length
*       17035      17134: gap of unknown length
*       17135      24051: contig of 6917 bp in length
*       24052      24151: gap of unknown length
*       24152      34072: contig of 9921 bp in length
*       34073      34172: gap of unknown length
*       34173      48742: contig of 14570 bp in length
*       48743      48842: gap of unknown length
*       48843      59843: contig of 11001 bp in length
*       59844      59943: gap of unknown length
*       59944      76882: contig of 16839 bp in length
*       76883      95694: gap of unknown length
*       95695      95794: gap of unknown length
*       95795      117154: contig of 21360 bp in length
*       117155     117254: gap of unknown length
*       117255     140616: contig of 23362 bp in length
*       140617     173491: gap of unknown length
*       140717     327755: contig of 32775 bp in length.

```

FEATURES

source

```

1..173491
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-241019"
3720..3819
/estimated_length=unknown
7607..7706
/estimated_length=unknown
10228..10327
/estimated_length=unknown
17035..17134
/estimated_length=unknown
24052..24151
/estimated_length=unknown
34073..34172
/estimated_length=unknown
48743..48842
/estimated_length=unknown
59844..59943
/estimated_length=unknown
76883..76882
/estimated_length=unknown
95695..95794
/estimated_length=unknown
117155..117254
/estimated_length=unknown
140617..140716
/estimated_length=unknown

```

ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 173491;
 Best Local Similarity 96.0%; Pred. No. 1.9e+02;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATCT 25
 Db 29112 AAAAAAAAAATCGCAACCAATCT 29136

RESULT 27

AC152354

LOCUS

DEFINITION

AC152354

Pan troglodytes chromosome X clone RP43-011D19 map human ortholog

q26.2, 2 ordered pieces.

AC152354

AC152354.1 GI:55251387

HTG: HTGS PHASE2; HTGS FULLTOP; HTGS_CANCELLED.

Pan troglodytes (chimpanzee)

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Pan.

1 (bases 1 to 189791)

Taudien,S., Taerner,S., Baumgart,C., Mueller,O. and Platzner,M.

DNA sequence of chimpanzee chromosome X

Unpublished

2 (bases 1 to 189791)

Lagemann,D. and Platzner,M.

Direct Submission

Submitted (03-NOV-2004) Genome Analysis, Institute of Molecular

Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany

Genome Center

Center: Institute of Molecular Biotechnology

Center code: IMB

Web site: <http://genome.imb-jena.de/>Contact: gacj-submit@genome.imb-jena.de

Project Information

Center project name: CX72

Center clone name: RP43-011D19

Summary Statistics

Sequencing clone based: PUC18; 86% of reads

Whole genome shotgun: 14% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 189654 bases at least Q40

Consensus quality: 189687 bases at least Q30

Consensus quality: 189691 bases at least Q20

Quality coverage: 10.15x

 This sequence was generated with the support of external whole genome shotgun data (<http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi>) as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

 Neighboring sequence information:
 This clone is overlapped by RP43-057B02.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

 * NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 58947: contig of 58947 bp in length
 * 58948 59047: gap of unknown length
 * 59048 189791: contig of 130744 bp in length.
 Location/Qualifiers

FEATURES

source
 1..189791
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="X"
 /map="human ortholog q26.2"
 /clone="RP43-011D19"
 175145..189691
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="RP43-057B02"
 /note="Overlapping clone"
 1..43
 /note="single stranded/single chemistry region"
 misc_feature
 143..214
 /note="single stranded/single chemistry region"
 misc_feature
 2451..2862
 /note="single stranded/single chemistry region"
 misc_feature
 2863..2960
 /note="single clone coverage"
 2961..3245
 /note="single stranded/single chemistry region"
 misc_feature
 4859..5057
 /note="single stranded/single chemistry region"
 misc_feature
 4949..4950
 /note="low quality region"
 15793..15878
 /note="single stranded/single chemistry region"
 misc_feature
 21939..21982
 /note="single stranded/single chemistry region"
 misc_feature
 23573..23746
 /note="single stranded/single chemistry region"
 misc_feature
 26635..26703
 /note="single stranded/single chemistry region"
 misc_feature
 28360..28437
 /note="single stranded/single chemistry region"
 misc_feature
 29168..29694
 /note="single stranded/single chemistry region"
 misc_feature
 42239..42910
 /note="single stranded/single chemistry region"
 misc_feature
 43030..43035
 /note="single stranded/single chemistry region"
 misc_feature
 43036..43073
 /note="single clone coverage"
 43074..43351
 /note="single stranded/single chemistry region"
 misc_feature
 55742..56092
 /note="single stranded/single chemistry region"
 misc_feature
 56290..56593
 /note="single stranded/single chemistry region"
 misc_feature
 58752..58773
 /note="single stranded/single chemistry region"
 misc_feature
 58830..59083
 /note="single stranded/single chemistry region"
 gap
 58948..59047
 /estimated_length=unknown
 58948
 /note="low quality region"
 59808..59861
 /note="single stranded/single chemistry region"

ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 189791;
 Best Local Similarity 96.0%; Pred. No. 1.8e+02;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 misc_feature
 87447..87851
 /note="single stranded/single chemistry region"
 misc_feature
 88748..88842
 /note="single stranded/single chemistry region"
 misc_feature
 111941..112084
 /note="single stranded/single chemistry region"
 misc_feature
 120720..120814
 /note="single stranded/single chemistry region"
 misc_feature
 134618..134658
 /note="single clone coverage"
 134618
 /note="low quality region"
 134724..134831
 /note="single stranded/single chemistry region"
 137592..137850
 /note="single stranded/single chemistry region"
 140932..141004
 /note="whole-genome shotgun only"
 144029..144060
 /note="single stranded/single chemistry region"
 16048..166217
 /note="single stranded/single chemistry region"

OR
 1
 DB 115168
 AAAAAAAAAATCGCAACCAATCT 25
 AAAAAAAAAATCGCAACCAATCT 115192

RESULT 28

AC068681/c

LOCUS Homo sapiens chromosome 6 clone RP11-671L6 map 6, WORKING DRAFT
 DEFINITION
 ACCESSION AC068681.2 GI:8316617
 VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 196831)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Baetsen, V., Beda, F.,
 Boguslavsky, L., Bouckgalter, B., Brown, A., Burdett, G.,
 Campoliano, A., Casale, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, D., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, D. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamasares, R., Landers, T., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marulis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, C., McPheters, R.,
 Meldrum, D., Menus, L., Milova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, T., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Plesant, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Sudramanatan, A., Talamas, J.,
 Teffaly, S., Theodore, D., Tirrell, A., Travers, M., Triggillo, J.,
 Vasiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

JOURNAL

TITLES

AUTHORS

TITLE

Direct Submission

JOURNAL
Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 196831)
REFERENCE
Birken, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Baatien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Caselle, A., Choepel, X., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Labocque, K., Lamarca, R., Landers, T., Lenoczky, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Menue, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olyar, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanil, C., Polara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu

Project Information
Center project name: L5642
Center clone name: 671_L8

Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187164 bases at least Q40
Consensus quality: 191936 bases at least Q30
Consensus quality: 193826 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 194931; sum-of-ctnigs
Quality coverage: 4.5 in Q20 bases; sum-of-ctnigs
Quality coverage: 4.4 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1018: contig of 1018 bp in length
* 1019 1018: gap of 100 bp
* 1119 1018: gap of 100 bp
* 2873: contig of 1755 bp in length
* 2874 2973: gap of 100 bp
* 2974 4656: contig of 1683 bp in length
* 4657 4756: gap of 100 bp
* 4757 7018: contig of 2262 bp in length
* 7019 7118: gap of 100 bp
* 7119 9644: contig of 2526 bp in length
* 9645 9744: gap of 100 bp
* 9745 11714: contig of 1970 bp in length
* 11715 11814: gap of 100 bp
* 11815 14827: contig of 3013 bp in length
* 14828 14927: gap of 100 bp

FEATURES
source
* 14928 17471: contig of 2544 bp in length
* 17472 17571: gap of 100 bp
* 17572 21911: contig of 4340 bp in length
* 21912 22011: gap of 100 bp
* 22012 27134: contig of 5123 bp in length
* 27135 32034: gap of 100 bp
* 32035 32134: gap of 100 bp
* 32135 39278: contig of 7144 bp in length
* 39279 39378: gap of 100 bp
* 39379 44125: contig of 4747 bp in length
* 44126 44225: gap of 100 bp
* 44226 58909: contig of 14684 bp in length
* 58910 59009: gap of 100 bp
* 59010 75495: contig of 16486 bp in length
* 75496 75595: gap of 100 bp
* 75596 95743: contig of 20148 bp in length
* 95744 95843: gap of 100 bp
* 95844 119614: contig of 23771 bp in length
* 119615 119714: gap of 100 bp
* 119715 143111: contig of 23397 bp in length
* 143112 143211: gap of 100 bp
* 143212 171502: contig of 28291 bp in length
* 171503 171603: gap of 100 bp
* 171603 196831: contig of 25229 bp in length.

Location/Qualifiers
1. .196831
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-671L8"
/clone_1lb="RPC1-11 Human Male BAC"
1. .1018
/note="assembly_fragment"
gap
misc_feature
1019. .1118
/estimated_length=100
misc_feature
1119. .2873
/note="assembly_fragment"
gap
misc_feature
2874. .2973
/estimated_length=100
misc_feature
2974. .4656
/note="assembly_fragment"
gap
misc_feature
4657. .4756
/estimated_length=100
misc_feature
4757. .7018
/note="assembly_fragment"
gap
misc_feature
7019. .7118
/estimated_length=100
misc_feature
7119. .9644
/note="assembly_fragment"
gap
misc_feature
9645. .9744
/estimated_length=100
misc_feature
9745. .11714
/note="assembly_fragment"
gap
misc_feature
11715. .11814
/estimated_length=100
misc_feature
11815. .14827
/note="assembly_fragment"
gap
misc_feature
14828. .14927
/estimated_length=100
misc_feature
14928. .17471
/note="assembly_fragment"
gap
misc_feature
17472. .17571
/estimated_length=100
misc_feature
17572. .21911
/note="assembly_fragment"
gap
misc_feature
21912. .22011
/estimated_length=100
misc_feature
22012. .27134
/note="assembly_fragment"
gap
misc_feature
27135. .27234

```
misc_feature      /estimated_length=100
                  /note="assembly_fragment"
gap              32035..32134
                  /estimated_length=100
misc_feature      32135..39278
                  /note="assembly_fragment"
gap              39279..39378
                  /estimated_length=100
misc_feature      39379..44125
                  /note="assembly_fragment"
gap              44126..44225
                  /estimated_length=100
misc_feature      44226..58909
                  /note="assembly_fragment"
                  clone_end:17

Query Match      93.6%; Score 23.4; DB 14; Length 196831;
Best Local Similarity 96.0%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY              1 AAAAAAAAAATCGCAACAAATCT 25
Db              169813 AAAAAAAAAATCGCAACAAATCT 169789

RESULT 29
AC016311
LOCUS           AC016311      224450 bp      DNA      linear      HTG 16-APR-2000
DEFINITION     Homo sapiens clone RP11-27H10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION      AC016311.3
VERSION        AC016311.3 GI:7577571
KEYWORDS       HTG; HTGS_PHASE0.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 224450)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-27H10
Unpublished
2 (bases 1 to 224450)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collimore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,W., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galegan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kam,L., Karatas,A., Klein,J.,
Lhoczy,J., Lieu,C., Locke,K., MacDonald,P., Margulis,N.,
McBarn,P., McGuirk,A., McKernan,K., McLaughlin,J., Meltrin,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vasilev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 16, 2000 this sequence version replaced gi:6970658.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W18R
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L496
Center clone name: 27_H_10
-----
```

* NOTE: This record contains 246 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```
1 728 827: contig of 727 bp in length
* 728 827: gap of 100 bp
* 828 1601: contig of 774 bp in length
* 1602 1701: gap of 100 bp
* 1702 2457: contig of 756 bp in length
* 2458 2557: gap of 100 bp
* 2558 3331: contig of 774 bp in length
* 3332 3432 3431: gap of 100 bp
* 3432 4213: contig of 782 bp in length
* 4214 4313: gap of 100 bp
* 4314 5079: contig of 766 bp in length
* 5080 5180 5179: gap of 100 bp
* 5180 5947: contig of 768 bp in length
* 5948 6047: gap of 100 bp
* 6048 6817: contig of 770 bp in length
* 6818 6917: gap of 100 bp
* 6918 7666: contig of 749 bp in length
* 7667 7766: gap of 100 bp
* 7767 8515: contig of 749 bp in length
* 8516 8615: gap of 100 bp
* 8616 9385: contig of 770 bp in length
* 9386 9485: gap of 100 bp
* 9486 10258: contig of 773 bp in length
* 10259 11338: contig of 780 bp in length
* 11339 11238: gap of 100 bp
* 11239 12014: contig of 776 bp in length
* 12015 12114: gap of 100 bp
* 12115 12876: contig of 762 bp in length
* 12877 12976: gap of 100 bp
* 12977 13757: contig of 781 bp in length
* 13758 13857: gap of 100 bp
* 13858 14618: contig of 761 bp in length
* 14619 14718: gap of 100 bp
* 14719 15475: contig of 757 bp in length
* 15476 15575: gap of 100 bp
* 15576 16341: contig of 766 bp in length
* 16342 16441: gap of 100 bp
* 16442 17185: contig of 744 bp in length
* 17186 17285: gap of 100 bp
* 17286 18056: contig of 771 bp in length
* 18057 18156: gap of 100 bp
* 18157 18932: contig of 776 bp in length
* 18933 19032: gap of 100 bp
* 19033 19798: contig of 766 bp in length
* 19799 19898: gap of 100 bp
* 19899 20670: contig of 772 bp in length
* 20671 20770: gap of 100 bp
* 20771 21545: contig of 775 bp in length
* 21546 21646: gap of 100 bp
* 21646 22400: contig of 755 bp in length
* 22401 22500: gap of 100 bp
* 22501 23268: contig of 768 bp in length
* 23269 24138: gap of 100 bp
* 24139 24238: gap of 100 bp
* 24239 25003: contig of 765 bp in length
* 25004 25103: gap of 100 bp
* 25104 25865: contig of 762 bp in length
* 25866 26725: gap of 100 bp
* 26726 26825: contig of 760 bp in length
```



```

* 26826 27599: contig of 774 bp in length
* 27600 27699: gap of 100 bp
* 27700 28500: contig of 801 bp in length
* 28501 28600: gap of 100 bp
* 28601 29383: contig of 783 bp in length
* 29384 30228: gap of 100 bp
* 30229 30328: gap of 100 bp
* 30329 31091: contig of 763 bp in length
* 31092 31191: gap of 100 bp
* 31192 31946: contig of 755 bp in length
* 31947 32046: gap of 100 bp
* 32047 32828: contig of 782 bp in length
* 32829 32928: gap of 100 bp
* 32929 33699: contig of 771 bp in length
* 33700 33799: gap of 100 bp
* 33800 34564: contig of 765 bp in length
* 34565 34664: gap of 100 bp
* 34665 35408: contig of 744 bp in length
* 35409 35508: gap of 100 bp
* 35509 36271: contig of 763 bp in length
* 36272 36371: gap of 100 bp
* 36372 37136: contig of 765 bp in length
* 37137 37236: gap of 100 bp
* 37237 38006: contig of 770 bp in length
* 38007 38106: gap of 100 bp
* 38107 38850: contig of 744 bp in length
* 38851 39733: contig of 783 bp in length
* 39734 39833: gap of 100 bp
* 39834 40619: contig of 786 bp in length
* 40620 40719: gap of 100 bp
* 40720 41471: contig of 752 bp in length
* 41472 41571: gap of 100 bp
* 41572 42342: contig of 771 bp in length
* 42343 42442: gap of 100 bp
* 42443 43207: contig of 765 bp in length
* 43208 43307: gap of 100 bp
* 43308 44050: contig of 743 bp in length
* 44051 44150: gap of 100 bp
* 44151 44910: contig of 760 bp in length
* 44911 45010: gap of 100 bp
* 45011 45791: contig of 781 bp in length
* 45792 45891: gap of 100 bp
* 45892 46650: contig of 759 bp in length
* 46651 46750: gap of 100 bp
* 46751 47511: contig of 761 bp in length
* 47512 47611: gap of 100 bp
* 47612 48339: contig of 728 bp in length
* 48340 48439: gap of 100 bp
* 48439 49195: contig of 756 bp in length
* 49196 49295: gap of 100 bp
* 49296 50069: contig of 774 bp in length
* 50070 50169: gap of 100 bp
* 50170 50916: contig of 747 bp in length
* 50917 51016: gap of 100 bp
* 51017 51770: contig of 754 bp in length
* 51771 51870: gap of 100 bp
* 51871 52633: contig of 763 bp in length
* 52634 52733: gap of 100 bp
* 52734 53486: contig of 753 bp in length
* 53487 53586: gap of 100 bp
* 53587 54330: contig of 744 bp in length
* 54331 54430: gap of 100 bp
* 54431 55122: contig of 692 bp in length
* 55123 55222: gap of 100 bp
* 55223 55974: contig of 752 bp in length
* 55975 56074: gap of 100 bp
* 56075 56844: contig of 770 bp in length
* 56845 57625: gap of 100 bp
* 57626 57725: contig of 681 bp in length
* 57726 58519: contig of 794 bp in length

```

```

* 58520 58619: gap of 100 bp
* 58620 59399: contig of 780 bp in length
* 59400 59499: gap of 100 bp
* 59500 60274: contig of 775 bp in length
* 60275 60374: gap of 100 bp
* 60375 61153: contig of 779 bp in length
* 61154 61253: gap of 100 bp
* 61254 62030: contig of 777 bp in length
* 62031 62130: gap of 100 bp

Query Match
Beet Local Similarity 93.6% Score 23.4; DB 14; Length 224450;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 64284 AAAAAAAAAATTCGCAACAAATCT 64308

RESULT 30
AC016311/C
LOCUS
DEFINITION Homo sapiens clone RP11-27H10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016311.3 GI:7577571
VERSION
KEYWORDS HTG; HTGS; PHASE0.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 224450)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-27H10
Unpublished
2 (bases 1 to 224450)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donlan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kam,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tittell,A., Vasilev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 16, 2000 this sequence version replaced gi:6970658.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4696
Center clone name: 27_H_10
-----
* NOTE: This record contains 246 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.

```

```
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 728      727: contig of 727 bp in length
* 828      827: gap of 100 bp
* 1602     1601: contig of 774 bp in length
* 1702     1701: gap of 100 bp
* 2458     2457: contig of 756 bp in length
* 2558     2557: gap of 100 bp
* 3332     3331: contig of 774 bp in length
* 3432     3431: gap of 100 bp
* 4214     4213: contig of 782 bp in length
* 4314     4313: gap of 100 bp
* 5080     5079: contig of 766 bp in length
* 5180     5179: gap of 100 bp
* 5948     5947: contig of 768 bp in length
* 6048     6047: gap of 100 bp
* 6818     6817: contig of 770 bp in length
* 6918     6917: gap of 100 bp
* 7667     7666: contig of 749 bp in length
* 7767     7766: gap of 100 bp
* 8516     8515: contig of 749 bp in length
* 8616     8615: gap of 100 bp
* 9386     9385: contig of 770 bp in length
* 9486     9485: gap of 100 bp
* 10258    10258: contig of 773 bp in length
* 10359    10358: gap of 100 bp
* 11339    11338: contig of 780 bp in length
* 11239    11238: gap of 100 bp
* 12014    12014: contig of 776 bp in length
* 12015    12114: gap of 100 bp
* 12115    12876: contig of 762 bp in length
* 12877    12976: gap of 100 bp
* 12977    13757: contig of 781 bp in length
* 13758    13857: gap of 100 bp
* 13858    14618: contig of 761 bp in length
* 14619    14718: gap of 100 bp
* 14719    15475: contig of 757 bp in length
* 15476    15575: gap of 100 bp
* 15575    16341: contig of 766 bp in length
* 1576     16441: gap of 100 bp
* 16342    17185: contig of 744 bp in length
* 16442    17285: gap of 100 bp
* 17286    18056: contig of 771 bp in length
* 18057    18156: gap of 100 bp
* 18157    18932: contig of 776 bp in length
* 18933    19032: gap of 100 bp
* 19033    19798: contig of 766 bp in length
* 19799    19898: gap of 100 bp
* 19899    20670: contig of 772 bp in length
* 20671    20770: gap of 100 bp
* 20771    21545: contig of 775 bp in length
* 21546    21645: gap of 100 bp
* 21646    22400: contig of 755 bp in length
* 22401    22500: gap of 100 bp
* 22501    23368: contig of 768 bp in length
* 23269    23368: gap of 100 bp
* 23369    24138: contig of 770 bp in length
* 24139    24238: gap of 100 bp
* 24239    25003: contig of 765 bp in length
* 25004    25103: gap of 100 bp
* 25104    25865: contig of 762 bp in length
* 25866    25965: gap of 100 bp
* 25966    26725: contig of 760 bp in length
* 26726    26825: gap of 100 bp
* 26826    27599: contig of 774 bp in length
* 27600    27699: gap of 100 bp
* 27700    28500: contig of 801 bp in length
* 28501    28600: gap of 100 bp
* 28601    29383: contig of 783 bp in length
* 29384    29483: gap of 100 bp
* 29484    30228: contig of 745 bp in length
*
*
* 30229    30328: gap of 100 bp
* 30329    31091: contig of 763 bp in length
* 31092    31191: gap of 100 bp
* 31192    31946: contig of 755 bp in length
* 31947    32046: gap of 100 bp
* 32047    32828: contig of 782 bp in length
* 32829    32928: gap of 100 bp
* 32929    33699: contig of 771 bp in length
* 33700    33799: gap of 100 bp
* 33800    34564: contig of 765 bp in length
* 34565    34664: gap of 100 bp
* 34665    35408: contig of 744 bp in length
* 35409    35508: gap of 100 bp
* 35509    36271: contig of 763 bp in length
* 36272    36371: gap of 100 bp
* 36372    37136: contig of 765 bp in length
* 37137    37236: gap of 100 bp
* 37237    38006: contig of 770 bp in length
* 38007    38106: gap of 100 bp
* 38107    38850: contig of 744 bp in length
* 38851    38950: gap of 100 bp
* 38951    39733: contig of 783 bp in length
* 39734    39833: gap of 100 bp
* 39834    40619: contig of 786 bp in length
* 40620    40719: gap of 100 bp
* 40720    41471: contig of 752 bp in length
* 41472    41571: gap of 100 bp
* 41572    42342: contig of 771 bp in length
* 42343    42442: gap of 100 bp
* 42443    43207: contig of 765 bp in length
* 43208    43308: gap of 100 bp
* 43308    44050: contig of 743 bp in length
* 44051    44150: gap of 100 bp
* 44151    44910: contig of 760 bp in length
* 44911    45010: gap of 100 bp
* 45011    45791: contig of 781 bp in length
* 45792    45891: gap of 100 bp
* 45892    46550: contig of 759 bp in length
* 46551    46750: gap of 100 bp
* 46751    47511: contig of 761 bp in length
* 47512    47611: gap of 100 bp
* 47612    48339: contig of 728 bp in length
* 48340    48439: gap of 100 bp
* 48440    49195: contig of 756 bp in length
* 49196    49295: gap of 100 bp
* 49296    50069: contig of 774 bp in length
* 50070    50169: gap of 100 bp
* 50170    50916: contig of 747 bp in length
* 50917    51016: gap of 100 bp
* 51017    51770: contig of 754 bp in length
* 51771    51870: gap of 100 bp
* 51871    52633: contig of 763 bp in length
* 52634    52733: gap of 100 bp
* 52734    53486: contig of 753 bp in length
* 53487    53586: gap of 100 bp
* 53587    54330: contig of 744 bp in length
* 54331    54430: gap of 100 bp
* 54431    55122: contig of 692 bp in length
* 55123    55222: gap of 100 bp
* 55223    55974: contig of 752 bp in length
* 55975    56074: gap of 100 bp
* 56075    56844: contig of 770 bp in length
* 56845    56944: gap of 100 bp
* 56945    57625: contig of 681 bp in length
* 57626    57725: gap of 100 bp
* 57726    58519: contig of 794 bp in length
* 58519    58619: gap of 100 bp
* 58620    59399: contig of 780 bp in length
* 59400    59499: gap of 100 bp
* 59500    60274: contig of 775 bp in length
* 60275    60374: gap of 100 bp
* 60375    61153: contig of 779 bp in length
* 61154    61253: gap of 100 bp
```

```

*      61254      62030: contig of 777 bp in length
*      62031      62130: gap of 100 bp

Query Match
Best Local Similarity 93.6%; Score 23.4; DB 14; Length 224450;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGCAACCAATCT 25
Db      194741 AAAAAAAAAATCGCAACCAATCT 194717

RESULT 31
LOCUS   BS000230      242704 bp      DNA      linear      PRI 12-JUN-2004
DEFINITION Pan troglodytes chromosome 22 clone:PTB-105H12, map 22, complete
sequences.
ACCESSION BS000230 BA000046
VERSION   BS000230.1 GI:37537497
KEYWORDS  HTG.
SOURCE    Pan troglodytes (chimpanzee)
ORGANISM  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Pan.

REFERENCE
AUTHORS   1
TITLE     The International Chimpanzee Chromosome 22 Consortium.
JOURNAL   DNA sequence and comparative analysis of chimpanzee chromosome 22
AUTHORS   Nature 429, 382-388 (2004)
TITLES    2 (bases 1 to 242704)
AUTHORS   Tsai, S., Liu, T., Wu, K., Liao, T. and Hsiao, K.
JOURNAL   Direct Submission
TITLES    Submitted (16-MAY-2003) Shih-Feng Tsai, National Health Research
AUTHORS   Institutes (NHRI), Division of Molecular and Genomic Medicine; 128,
JOURNAL   Yen-Chiu-Yuan Road, Sec 2, Taipei 115, Taiwan
TITLES    (E-mail: petersai@nhri.org.tw, URL: http://www.nhri.org.tw/,
AUTHORS   Tel: 886-2-28267319, Fax: 886-2-28200552)
JOURNAL   The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
TITLES    *Chinese National Human Genome Center at Shanghai, Shanghai, China;
AUTHORS    *GBR, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
JOURNAL    of Molecular Biotechnology, Jena, Germany; *KRIIBB Genome Research
TITLES    Center, Daejeon, Korea;
AUTHORS    *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
JOURNAL    *National Institute of Genetics, Mishima, Japan;
TITLES    *National Yang Ming University Genome Research Center, Taipei,
AUTHORS    Taiwan;
JOURNAL    *RIKEN Genomic Sciences Center, Yokohama, Japan.
TITLES    ----- Genome Center
AUTHORS    Center: National Yang Ming University Genome Research Center
JOURNAL    code: YMG
TITLES    Web site: http://genome.ym.edu.tw/
AUTHORS    Contact: sequence@ym.edu.tw
JOURNAL    ----- Project Information
TITLES    Center project name: The Chimpanzee Chromosome 22 Sequencing Project
AUTHORS    Center clone name: HC
JOURNAL    ----- Summary Statistics
TITLES    Sequencing vector: pUC18; 100% of reads
AUTHORS    Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly
JOURNAL    Program: Phrap; version 0.990319
TITLES    Consensus quality: 242,677 bases at least Q40
AUTHORS    Consensus quality: 242,704 bases at least Q30
JOURNAL    Consensus quality: 242,704 bases at least Q20
TITLES    -----
AUTHORS    This sequence was finished as follows unless otherwise noted: all
JOURNAL    regions were double stranded, sequenced with an alternate
REFERENCE    chemistry, or covered by high quality data (i.e., phred quality >=
AUTHORS    30).
JOURNAL    an attempt was made to resolve all sequencing problems, such as
REFERENCE    compressions and repeats; all regions were covered by at one
AUTHORS    subclone or more than one M13 subclone;
JOURNAL    and the assembly was confirmed by restriction digest.
TITLES    -----

```

```

Source information:
The PTB1 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Aseo Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKS145

-----
Sequence Quality Assessment:
This entry has been annotated with sequence
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in
10,000 bp.
-----
Neighboring clones: RP43-015P20(left) and PTB-071B01(right).

FEATURES
source
location/Qualifiers
1..242704
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-105H12"
/clone_lib="PTB1 chimpanzee BAC"

ORIGIN
Query Match
Best Local Similarity 93.6%; Score 23.4; DB 8; Length 242704;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGCAACCAATCT 25
Db      145859 AAAAAAAAAATCGCAACCAATCT 145883

RESULT 32
LOCUS   AP001747      340000 bp      DNA      linear      PRI 21-MAY-2003
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 91/105.
ACCESSION AP001747 AL163292 BA000005
VERSION   AP001747.1 GI:7768768
KEYWORDS
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.

REFERENCE
AUTHORS   1
Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T.,
Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E.,
Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K.,
Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R.,
Patterson, D., Reichwald, K., Rump, A., Schlichthabel, M., Schudy, A.,
Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K.,
Asakawa, S., Shintani, A., Sasaki, K., Nagamine, K., Mitsuyama, S.,
Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordstreck, G.,
Hornischer, K., Barand, P., Scharte, M., Schoen, O., Desario, A.,
Reichelt, J., Kauer, G., Blocker, H., Ramser, J., Beck, A., Klages, S.,
Hennig, S., Rieselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K.,
Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and
Yaspo, M. L.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

```

TITLE
JOURNAL

Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Risseltmann,L., Dagnid,B., Wehrmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and Yaspo,M.L.

COMMENT

Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * Gbf, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
On May 30, 2000 this sequence version replaced gi:7717403.
The chromosome 21 mapping and sequencing consortium consisting of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamiara 228-8555, Japan,
* e.mail: hatorigsec.riken.go.jp
* URL: http://hpg.gsc.riken.go.jp/

and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-studmit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/

and
* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,
* e.mail: nahimizu@dmh-med.keio.ac.jp
* URL: http://www.dmh.med.keio.ac.jp/

and
* Gbf, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/

and
* Max-Planck Institute for Molecular Genetics,
* Innestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/

FEATURES

AL163292: Submitted (10-Apr-2000).

Location/Qualifiers

source

1. .340000

source

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

<1. .2534

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="KB99468, 5' partial"

/clone_lib="Keio BAC library"

/note="Accession No. AP001624"

<1. .106779

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="KB907F12, 5' partial"

/clone_lib="Keio BAC library"

/note="Accession No. AP001625"

66930. .214714

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="KB159F8"

/clone_lib="Keio BAC library"

/note="Accession No. AP001626"

166876. .292251

source

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="KB16188"

/clone_lib="Keio BAC library"

/note="Accession No. AP001627"

238164. .3940000

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.3"

/clone="KB51A8, 3' partial"

/clone_lib="Keio BAC library"

/note="Accession No. AP001628"

complement (1371. .1484)

/note="L2"

/rpt_family="LINE/L2"

/rpt_type=DISPERSED

complement (1485. .1644)

/note="MIR"

/rpt_family="SINE/MIR"

/rpt_type=DISPERSED

complement (1744. .2040)

/note="AluX"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

complement (2389. .2806)

/note="MSTRB"

/rpt_family="LTR/MaLR"

/rpt_type=DISPERSED

complement (2981. .3286)

/note="AluX"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

complement (3920. .4158)

/note="HAI1"

/rpt_family="LINE/Other"

/rpt_type=DISPERSED

4455. .4516

/note="L2"

/rpt_family="LINE/L2"

/rpt_type=DISPERSED

4911. .4931

/note="L1M1"

/rpt_family="LINE/L1"

/rpt_type=DISPERSED

complement (4932. .5183)

/note="AluSc"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

5184. .6211

/note="L1M1"

/rpt_family="LINE/L1"

/rpt_type=DISPERSED

6212. .6361

/note="L1MA9"

/rpt_family="LINE/L1"

/rpt_type=DISPERSED

6362. .6682

/note="AluDo"

/rpt_family="SINE/Alu"

/rpt_type=DISPERSED

6683. .7078

/note="L1MA9"

/rpt_family="LINE/L1"

/rpt_type=DISPERSED

complement (7184. .7373)

/note="L2"

/rpt_family="LINE/L2"

/rpt_type=DISPERSED

```

misc_feature      8402..8990
                    /note="CpG island"
repeat_region     complement(10533..10573)
                    /note="L1PA13"
                    /rpt_family="LINE/L1"
                    /rpt_type=DISPERSED
repeat_region     complement(10620..10916)
                    /note="AluX"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
                    complement(11979..12150)
                    /note="MIR"
                    /rpt_family="SINE/MIR"
                    /rpt_type=DISPERSED
                    complement(13952..14225)
                    /note="MER34"
                    /rpt_family="LTR/MER21-group"
                    /rpt_type=DISPERSED

repeat_region     93.6%; Score 23.4; DB 8; Length 340000;
                    Best Local Similarity 96.0%; Pred. No. 1.5e+02;
                    Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCGAACAATCT 25
Db      330896 AAAAAAAAAATCGAACAATCT 330920

RESULT 33
AC092209/c  AC092209      14026 bp      DNA      linear      PRI 08-OCT-2003
DEFINITION Homo sapiens BAC clone RP11-592C12 from 7, complete sequence.
VERSION     AC092209.5   GI:15638886
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
            1 (bases 1 to 14026)
            Suleston, J.E. and Wilson, R.
            Toward a complete human genome sequence
            JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
            PUBMED 9847074
            2 (bases 1 to 14026)
            Cotton, M., Meyer, R. and Waligorski, J.
            The sequence of Homo sapiens BAC clone RP11-592C12
            JOURNAL Unpublished (2001)
            3 (bases 1 to 14026)
            Waterston, R.H.
            Direct Submissio
            REFERENCE JOURNAL
            4 (bases 1 to 14026)
            Submitted (28-JUN-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            5 (bases 1 to 14026)
            Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            6 (bases 1 to 14026)
            Direct Submissio
            REFERENCE JOURNAL
            7 (bases 1 to 14026)
            Submitted (08-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Sep 18, 2001 this sequence version replaced gi:1514355.
            COMMENT ----- Genome Center

```

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: saplens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0592C12

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:
 The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Caranese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-52501, 2000 bp overlap the clone sequenced to the right is A093117, 2000 bp overlap. Actual start of this clone is at base position 72833 of RP11-52501.
 Location/Qualifiers

```

FEATURES
source      1..14026
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="7"
            /clone="RP11-592C12"
            /clone_lib="RPc1-11"
            /size="380..356"
            /note="CpG island (%GC=62.7, o/e=0.76, #CpGs=22)"
            /rpt_family="MER1_type"
repeat_region     380..530
                    /rpt_family="MER1_type"
                    566..716
                    /rpt_family="MIR"
                    1209..1251
                    /rpt_family="TG(n)"
                    2947..3076
                    /rpt_family="MaLR"
                    3101..4092
                    /rpt_family="L2"
                    4376..4538
                    /rpt_family="MIR"
                    5081..5126
                    /rpt_family="L1"
                    5127..5423
                    /rpt_family="Alu"
                    5424..5615
repeat_region

```

```

repeat_region      /rpt_family="L1"
                    5803..5913
repeat_region      /rpt_family="L1"
                    6053..6187
repeat_region      /rpt_family="L1"
                    6188..6496
repeat_region      /rpt_family="Alu"
                    6497..6577
repeat_region      /rpt_family="L1"
                    6648..7188
repeat_region      /rpt_family="ERV"
                    8805..9163
repeat_region      /rpt_family="MALR"
                    9206..9559
repeat_region      /rpt_family="MALR"
                    9551..9689
repeat_region      /rpt_family="MERL_type"
                    10523..10591
repeat_region      /rpt_family="MIR"
                    11279..11495
repeat_region      /rpt_family="MERL_type"
                    12457..12507
repeat_region      /rpt_family="MIR"
                    12508..12766
repeat_region      /rpt_family="Alu"
                    12767..12957
repeat_region      /rpt_family="MIR"
                    13526..13626
repeat_region      /rpt_family="MIR"
                    13728..14026
repeat_region      /rpt_family="MALR"

```

```

Query Match      89.6%; Score 22.4; DB 8; Length 14026;
Best Local Similarity 95.8%; Pred. No. 8.2e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 2 AAAAAAAAAATGCAACAACTCT 25
Db 11398 AAAAAAAAAATGCAACAACTCT 11375

RESULT 34
AC016014/C      68419 bp      DNA      linear      HTG 13-JUL-2000
LOCUS      AC016014
DEFINITION      Homo sapiens clone RP11-26J10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION      AC016014.2 GI:9134660
VERSION      HTG; HTGS_PHASE0.
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 68419)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-26J10
Unpublished
2 (bases 1 to 68419)

```

```

REFERENCE
AUTHORS      Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baldin,U., Barina,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
              Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
              Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
              Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Johnson,R., Jones,C., Kam,L., Karasik,A., Klein,J.,
              Lehotzky,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N.,
              McKernan,P., McGurt,A., McKernan,K., McLaughlin,J., Meidrich,J.,
              Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Teslaye,S., Tirrell,A., Vasilev,H., Vo,A., Wheeler,J., Wu,X.,
              Wymann,D., Ye,W.J., Zimmer,A. and Zody,M.

```

TITLE JOURNAL

COMMENT

```

Direct Submission
Submitted (18-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6449522.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4643
Center clone name: 26_J_10
-----
* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
788: contig of 788 bp in length
789
888: gap of 100 bp
889
1666: contig of 778 bp in length
1667
1766: gap of 100 bp
1767
2567: contig of 801 bp in length
2568
2667: gap of 100 bp
2668
3636: contig of 968 bp in length
3637
3736: gap of 100 bp
3737
4595: contig of 859 bp in length
4596
4695: gap of 100 bp
4696
5517: contig of 822 bp in length
5518
5617: gap of 100 bp
5618
6436: contig of 819 bp in length
6437
6536: gap of 100 bp
6537
7361: contig of 825 bp in length
7362
7461: gap of 100 bp
7462
8283: contig of 822 bp in length
8284
8383: gap of 100 bp
8384
9171: contig of 788 bp in length
9172
9271: gap of 100 bp
9272
10074: contig of 803 bp in length
10075
10174: gap of 100 bp
10175
10962: contig of 788 bp in length
10963
11062: gap of 100 bp
11063
11845: contig of 783 bp in length
11846
11945: gap of 100 bp
11946
12758: contig of 813 bp in length
12759
12858: gap of 100 bp
12859
13619: contig of 761 bp in length
13620
13719: gap of 100 bp
13720
14542: contig of 823 bp in length
14543
14642: gap of 100 bp
14643
15428: contig of 786 bp in length
15429
15528: gap of 100 bp
15529
16328: contig of 800 bp in length
16329
16428: gap of 100 bp
16429
17221: contig of 793 bp in length
17222
17321: gap of 100 bp
17322
18099: contig of 778 bp in length
18100
18199: gap of 100 bp
18200
18980: contig of 781 bp in length
18981
19080: gap of 100 bp
19081
19869: contig of 789 bp in length
19870
19970: gap of 100 bp
19971
20807: contig of 838 bp in length
20808
20907: gap of 100 bp

```

```

* 20908 21769: contig of 862 bp in length
* 21770 21869: gap of 100 bp
* 21870 22649: contig of 780 bp in length
* 22650 22749: gap of 100 bp
* 22750 23545: contig of 796 bp in length
* 23546 23645: gap of 100 bp
* 23646 24442: contig of 797 bp in length
* 24443 24542: gap of 100 bp
* 24543 25345: contig of 803 bp in length
* 25346 25445: gap of 100 bp
* 25446 26237: contig of 782 bp in length
* 26238 26328: gap of 100 bp
* 26328 27126: contig of 793 bp in length
* 27127 27226: gap of 100 bp
* 27227 28039: contig of 813 bp in length
* 28040 28139: gap of 100 bp
* 28140 28965: contig of 826 bp in length
* 28966 29066: gap of 100 bp
* 29066 29854: contig of 789 bp in length
* 29855 29954: gap of 100 bp
* 29955 30736: contig of 782 bp in length
* 30737 30836: gap of 100 bp
* 30837 31593: contig of 757 bp in length
* 31594 31693: gap of 100 bp
* 31694 32483: contig of 790 bp in length
* 32484 32583: gap of 100 bp
* 32584 33391: contig of 808 bp in length
* 33392 33491: gap of 100 bp
* 33492 34285: contig of 794 bp in length
* 34286 34385: gap of 100 bp
* 34386 35223: contig of 838 bp in length
* 35224 35323: gap of 100 bp
* 35324 36128: contig of 805 bp in length
* 36129 36228: gap of 100 bp
* 36229 37039: contig of 811 bp in length
* 37040 37139: gap of 100 bp
* 37140 37899: contig of 760 bp in length
* 37900 37999: gap of 100 bp
* 38000 38779: contig of 780 bp in length
* 38780 38879: gap of 100 bp
* 38880 39673: contig of 794 bp in length
* 39674 39773: gap of 100 bp
* 39774 40594: contig of 821 bp in length
* 40595 40694: gap of 100 bp
* 40695 41475: contig of 781 bp in length
* 41476 41575: gap of 100 bp
* 41576 42363: contig of 788 bp in length
* 42364 42463: gap of 100 bp
* 42464 43252: contig of 789 bp in length
* 43253 43352: gap of 100 bp
* 43353 44179: contig of 827 bp in length
* 44180 44279: gap of 100 bp
* 44280 45104: contig of 825 bp in length
* 45105 45205: gap of 100 bp
* 45205 46034: contig of 830 bp in length
* 46035 46134: gap of 100 bp
* 46135 46936: contig of 802 bp in length
* 46937 47036: gap of 100 bp
* 47037 47869: contig of 833 bp in length
* 47870 47969: gap of 100 bp
* 47970 48770: contig of 801 bp in length
* 48771 48870: gap of 100 bp
* 48871 49668: contig of 798 bp in length
* 49669 49768: gap of 100 bp
* 49769 50560: contig of 792 bp in length
* 50561 50660: gap of 100 bp
* 50661 51418: contig of 758 bp in length
* 51419 51518: gap of 100 bp
* 51519 52301: contig of 783 bp in length
* 52302 52401: gap of 100 bp
* 52402 53189: contig of 788 bp in length
* 53190 53289: gap of 100 bp
* 53290 54081: contig of 792 bp in length

```

```

* 54082 54181: gap of 100 bp
* 54182 54984: contig of 803 bp in length
* 54985 55084: gap of 100 bp
* 55085 55863: contig of 779 bp in length
* 55864 55963: gap of 100 bp
* 55964 56778: contig of 815 bp in length
* 56779 56878: gap of 100 bp
* 56879 57657: contig of 779 bp in length
* 57658 57757: gap of 100 bp
* 57758 58554: contig of 797 bp in length
* 58555 58654: gap of 100 bp
* 58655 59472: contig of 818 bp in length
* 59473 59572: gap of 100 bp
* 59573 60360: contig of 788 bp in length
* 60361 60460: gap of 100 bp
* 60461 61240: contig of 780 bp in length
* 61241 61340: gap of 100 bp
* 61341 62195: contig of 855 bp in length
* 62196 62295: gap of 100 bp
* 62296 63086: contig of 791 bp in length
* 63087 63186: gap of 100 bp
* 63187 63978: contig of 792 bp in length
* 63979 64078: gap of 100 bp
* 64079 64874: contig of 796 bp in length
* 64875 64974: gap of 100 bp

Query Match      89.6%; Score 22.4; DB 14; length 68419;
Best Local Similarity 95.8%; Pred. No. 5.1e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCGCAACAAATC 24
Db      20351 AAAAAAAAAATCGCAAAAAAAAAATC 20328

RESULT 35
LOCUS      AC113340
DEFINITION Homo sapiens chromosome 5 clone CTRB-164L20, complete sequence.
ACCESSION  AC113340
VERSION    AC113340.2 GI:19747174
KEYWORDS   HTG.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE  1 (bases 1 to 84791)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Unpublished
            2 (bases 1 to 84791)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (01-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            3 (bases 1 to 84791)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            Direct Submission
            Submitted (27-MAR-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            On Mar 27, 2002 this sequence version replaced gi:19033454.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.sbgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.4.
            Location/Qualifiers
                1..84791
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
FEATURES
    source

```

ORIGIN /chromosome="5"
/clone="CTB-164L20"

Query Match 89.6%; Score 22.4; DB 8; Length 84791;
Best Local Similarity 95.8%; Pred. No. 4.8e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATC 24
Db 9667 AAAAAAAAAATCGCAACAAATC 9690

RESULT 36
AC011385 134599 bp DNA linear PRI 30-APR-2002
DEFINITION Homo sapiens chromosome 5 clone CTB-15L17, complete sequence.
AC011385
VERSION AC011385.6 GI:20340444
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE 1 (bases 1 to 134599)
DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL
TITLE Direct Submission
UNpublished
2 (bases 1 to 134599)
DOE Joint Genome Institute.

REFERENCE 2 (bases 1 to 134599)
Direct Submission
Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134599)
DOE Joint Genome Institute.

REFERENCE 4 (bases 1 to 134599)
Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE 5 (bases 1 to 134599)
Submitted (30-APR-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Apr 30, 2002 this sequence version replaced gi:19224747.

COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40.99.9% of Sequence;
Estimated Total Number of Errors is 0.1.

FEATURES
source
1.134599
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-15L17"

ORIGIN
Query Match 89.6%; Score 22.4; DB 8; Length 134599;
Best Local Similarity 95.8%; Pred. No. 4.2e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATC 24
Db 4762 AAAAAAAAAATCGCAACAAATC 4785

RESULT 37
AC142718 147263 bp DNA linear HTG 09-APR-2003
AC142718/c
LOCUS
DEFINITION Macaca mulatta clone CH250-272R21, *** SEQUENCING IN PROGRESS ***.

AC0142718
HTG; HTGS; PHASE2; HTGS_PGI.
Macaca mulatta (rhesus monkey)
Macaca mulatta

REFERENCE 1 (bases 1 to 147263)
Genus, M. and Milosavljevic, A.
Pooled genomic indexing (PGI): mathematical analysis and experiment
design
(in) Guigo, R. and Gusfield, D. (Eds.);
ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;
Springer (2002)

JOURNAL
TITLE Direct Submission
UNpublished
2 (bases 1 to 147263)
Submitted (05-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

AC142718.1 GI:29567357
HTG; HTGS; PHASE2; HTGS_PGI.
Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecoidea; Macaca.
1 (bases 1 to 147263)
Genus, M. and Milosavljevic, A.
Pooled genomic indexing (PGI): mathematical analysis and experiment
design
(in) Guigo, R. and Gusfield, D. (Eds.);
ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;
Springer (2002)
2 (bases 1 to 147263)
Submitted (05-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 147263)
Submitted (09-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: LCVN

Center clone name: CH250-272N21

----- Summary Statistics

Chemistry: Dye-terminator Big Dye; inf% of reads

Consensus quality: 1956 bases at least Q40

Consensus quality: 5110 bases at least Q30

Consensus quality: 6437 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

* NOTE: The contigs are based on the application

* of the PCI method using the Human genome (NCBI build 31)

* as the comparative genome.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

1 147263: contig of 147263 bp in length.

Location/Qualifiers

1. 147263

/organism="Macaca mulatta"

/mol_type="genomic DNA"

/db_xref="taxon:9544"

/clone="CH250-272N21"

1. 147263

/note="assembly_name:CH250-272N21.1A"

/note="assembly_name:CH250-272N21.1A"

CONFIDENCE: 0.83"

ORIGIN

Query Match 89.6%; Score 22.4; DB 14; Length 147263;

Best Local Similarity 95.8%; Pred. No. 4.1e+02;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATC 24

Db 609 AAAAAAAAAATCTCAACAAATC 586

RESULT 38

CR936486 152010 bp DNA linear HTG 16-JUL-2005

LOCUS CR936486

DEFINITION Danio rerio chromosome 5 clone DKEX-279C17, WORKING DRAFT SEQUENCE,

2 unordered pieces.

ACCESSION CR936486

VERSION CR936486.4 GI:70910571

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 152010)

McLaren,S.

Direct Submission

Submitted (14-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

fish-help@sanger.ac.uk Clone requests:

http://www.sanger.ac.uk/projects/D_rerio/fags.shtml#dataeight

On Jul 16, 2005 this sequence version replaced gi:6098137.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: fish-help@sanger.ac.uk

----- Project Information

Center project name: zK279C17

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 151252 bases at least Q40

Consensus quality: 151363 bases at least Q30

Consensus quality: 151549 bases at least Q20

Insert size: 151310; sum-of-contigs

Insert size: 108033; 43.1% error; agarose-fp

Quality coverage: 8.72x in Q20 bases; sum-of-contigs Quality

coverage: 12.45x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 19099: contig of 19099 bp in length

* 19100 19199: gap of 100 bp

* 19200 152010: contig of 132811 bp in length.

Location/Qualifiers

1. 152010

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/chromosome="5"

/clone="DKEX-279C17"

1. 19099

/note="assembly_fragment:00013"

19200. 152010

/note="assembly_fragment:00286.0"

ORIGIN

Query Match 89.6%; Score 22.4; DB 14; Length 152010;

Best Local Similarity 95.8%; Pred. No. 4e+02;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGCAACAAATCT 25

Db 98927 AAAAAAAAAATCGCAACAAATCT 98950

RESULT 39

AC144197 162753 bp DNA linear HTG 09-APR-2003

LOCUS AC144197

DEFINITION Macaca mulatta clone CH250-270H21, *** SEQUENCING IN PROGRESS ***.

ACCESSION AC144197

VERSION AC144197.1 GI:29649887

KEYWORDS HTG; HTGS PHASE2; HTGS_PGI.

SOURCE Macaca mulatta (rhesus monkey)

ORGANISM Macaca mulatta

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Cercopithecoidea; Cercopithecinae; Macaca.

1 (bases 1 to 162753)

Cauro,M. and Milosavljevic,A.

Pooled genomic indexing (PGI): mathematical analysis and experiment

design (in) Guigo,R. and Gusfield,D. (Eds.);

ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI

2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;

Springer (2002)

2 (bases 1 to 162753)

Milosavljevic,A., Sodergren,E., Cauro,M., Li,B., Jackson,A.R.,

Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Aisbrook,S.L.,

```

Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J.,
Benton,J., Birmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bulay,C.,
Butch,P., Butkelt,C., Butrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Egan,A., Eamhart,C., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gables,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gottrell,J.H., Guevara,W.,
Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsl,F.,
Howard,S., Huber,J., Huijck,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudah,S.,
Karlsen,E., Kelly,S., Khan,U., King,L., Korvab,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W.,
Loudesed,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Marondei,I., Martin,R.,
Martindale,A., Martinez,E., Massey,B., Mawhney,E., McLeod,M.P.,
Meador,M., Mei,G., Mercher,S., Metzker,M., Miller,A., Miner,G.,
Morris,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,B., Nwokenwo,S.,
Ogund,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Plimue,A., Pu,L.L.,
Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shih,C.,
Shoshitari,N., Sisson,I., Sodergren,R., Sonaike,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabors,P., Tameria,A.,
Tameria,K., Tang,H., Tansley,D., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D.,
Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watlington,S., Williams,G., Williamson,A.,
Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorilla,S., Zuccheri-Lapati,R., Weinstein,G., and Gibbs,R.
Direct Submission
Unpublished
3 (bases 1 to 162753)
Worley,K.C.
Direct Submission
Submitted (09-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: CH250-270H21
Center clone name: CH250-270H21
----- Summary Statistics
Chemistry: Dye-terminator Big Dye: inf of reads
Chemistry: Dye-terminator Big Dye: inf of reads
Consensus quality: 22987 bases at least Q40
Consensus quality: 25411 bases at least Q30
Consensus quality: 28409 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: The contigs are based on the application
* of the PGI method using the Human genome (NCBI build 31)
* as the comparative genome.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have

```

```

FEATURES      * 124707 163408: contig of 38702 bp in length.
source
1.163408
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7q"
gap
31082..31181
/estimated_length=unknown
gap
57358..57457
/estimated_length=unknown
gap
82133..82232
/estimated_length=unknown
gap
89982..90081
/estimated_length=unknown
gap
95189..95288
/estimated_length=unknown
gap
99395..99494
/estimated_length=unknown
gap
103256..103355
/estimated_length=unknown
gap
111104..111203
/estimated_length=unknown
gap
115964..116063
/estimated_length=unknown
gap
124607..124706
/estimated_length=unknown

ORIGIN
Query Match      89.6%; Score 22.4; DB 14; Length 163408;
Best Local Similarity 95.8%; Pred. No. 3.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGCAAAACAATCT 25
    |||||
Db 76830 AAAAAAAAAATCGCAAAACAATCT 76807

RESULT 41
AP002392 169212 bp DNA linear PRI 15-MAR-2003
LOCUS Homo sapiens genomic DNA, chromosome 11q clone:RP11-691F15,
DEFINITION complete sequences.
ACCESSION AP002392
VERSION AP002392.3 GI:14189773
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 Hattori,M., Iehli,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Homo sapiens genomic DNA
Published Only in Database (2000)
2 (bases 1 to 169212)
Hattori,M., Iehli,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Direct Submision
Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On May 22, 2001 this sequence version replaced gi:9857604.

COMMENT
FEATURES
source
1.169212
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"

```

```

ORIGIN
/clone="RP11-691F15"

Query Match      89.6%; Score 22.4; DB 8; Length 169212;
Best Local Similarity 95.8%; Pred. No. 3.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGCAAAACAATCT 25
    |||||
Db 52593 AAAAAAAAAATCGCAAAACAATCT 52570

RESULT 42
AC073645 178088 bp DNA linear HTG 15-JUL-2000
LOCUS Homo sapiens chromosome 11 clone RP11-539G23, WORKING DRAFT
DEFINITION SEQUENCE, 23 unordered pieces.
AC073645
AC073645.2 GI:8959169
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 178088)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 178088)
Waterston,R.H.
Direct Submission
Submitted (27-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8748935.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0539G23
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161940 bases at least Q40
Consensus quality: 166459 bases at least Q30
Insert size: 18300; agarose-fp
Insert size: 175888; sum-of-contigs
Quality coverage: 3.79 in Q20 bases; agarose-fp
Quality coverage: 4.03 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1282: contig of 1282 bp in length
* 1283 1382: gap of unknown length
* 1383 3347: contig of 1965 bp in length
* 3348 3447: gap of unknown length
* 3448 5219: contig of 1772 bp in length
* 5220 5320: gap of unknown length
* 5320 6553: contig of 1234 bp in length
* 6553 6653: gap of unknown length

```

*	6654	8477:	contig of 1824 bp in length
*	8478	8577:	gap of unknown length
*	8578	11196:	contig of 2619 bp in length
*	11197	11296:	gap of unknown length
*	11297	14978:	contig of 3682 bp in length
*	14979	15078:	gap of unknown length
*	15079	19404:	contig of 4326 bp in length
*	19405	19504:	gap of unknown length
*	19505	23316:	contig of 3812 bp in length
*	23317	23416:	gap of unknown length
*	23417	27548:	contig of 4132 bp in length
*	27549	27648:	gap of unknown length
*	27649	33455:	contig of 5807 bp in length
*	33456	33555:	gap of unknown length
*	33556	39328:	contig of 5773 bp in length
*	39329	39428:	gap of unknown length
*	39429	47444:	contig of 8016 bp in length
*	47445	47544:	gap of unknown length
*	47545	54494:	contig of 6550 bp in length
*	54495	54594:	gap of unknown length
*	54595	65092:	contig of 10498 bp in length
*	65093	65192:	gap of unknown length
*	65193	75930:	contig of 10738 bp in length
*	75931	76030:	gap of unknown length
*	76031	89491:	contig of 13461 bp in length
*	89492	89591:	gap of unknown length
*	89592	99513:	contig of 9922 bp in length
*	99514	99613:	gap of unknown length
*	99614	112379:	contig of 12766 bp in length
*	112380	112479:	gap of unknown length
*	112480	122229:	contig of 9750 bp in length
*	122230	122329:	gap of unknown length
*	122330	138387:	contig of 16058 bp in length
*	138388	138487:	gap of unknown length
*	138488	153753:	contig of 15266 bp in length
*	153754	153853:	gap of unknown length
*	153854	178088:	contig of 24235 bp in length

FEATURES
SOURCE

misc_feature	1.
gap	/note="assembly_name:Contig11" 1282. 1283. estimated_length=unknown 1381. 1347
misc_feature	/note="assembly_name:contig12" 3346. 3447 estimated_length=unknown 3448. 5219
misc_feature	/note="assembly_name: contig14 " 5220. 5319 estimated_length=unknown 5320. 6553
misc_feature	/note="assembly_name:Contig15 " 6554. 6653 estimated_length=unknown 6654. 8477
misc_feature	/note="assembly_name: contig16 " 8478. 8577 estimated_length=unknown 8578. 11156
misc_feature	/note="assembly_name:Contig17 " 11197. 11296 estimated_length=unknown 11297. 14578
misc_feature	/note="assembly_name: Contig18 " 14979. 15078 estimated_length=unknown 15079. 19104

	misc_feature	19505..2316	/note="assembly_name:Contig19"
	gap	19405..19504	clone_end:SP6
	gap	19405..19504	vector_side:right"
	misc_feature	19505..2316	/estimated_length=unknown
	gap	23317..23416	/note="assembly_name:Contig20"
	misc_feature	23417..27548	/estimated_length=unknown
	gap	27549..27648	/note="assembly_name:Contig21"
	misc_feature	27649..33455	/estimated_length=unknown
	gap	33456..33555	/note="assembly_name:Contig22"
	misc_feature	33556..39328	/estimated_length=unknown
	gap	39329..39428	/note="assembly_name:Contig23"
	misc_feature	39429..47444	/estimated_length=unknown
	gap	47445..47544	/note="assembly_name:Contig24"
	misc_feature	47545..54494	/estimated_length=unknown
	gap	54495..54594	/note="assembly_name:Contig25"
	misc_feature	54595..65092	/estimated_length=unknown
	gap	65093..65192	/note="assembly_name:Contig26"
	misc_feature	65193..75930	/estimated_length=unknown
	gap	75931..76030	/note="assembly_name:Contig27"
	misc_feature	76031..89491	/estimated_length=unknown
	gap	89492..89591	/note="assembly_name:Contig28"
	misc_feature	89592..99513	/estimated_length=unknown
	gap	99514..99613	/note="assembly_name:Contig29"
	misc_feature	99614..112379	/estimated_length=unknown
	gap	112380..112479	/note="assembly_name:Contig30"
	misc_feature	112480..122229	/estimated_length=unknown
	gap	122230..122329	/note="assembly_name:Contig31"
	misc_feature	122330..138387	/estimated_length=unknown
	gap	138388..138487	/note="assembly_name:Contig32"
	misc_feature	138488..153753	/estimated_length=unknown
	gap	153754..153853	/note="assembly_name:Contig33"
	misc_feature	153854..178088	/estimated_length=unknown
	gap	178089..178088	/note="assembly_name:Contig34"

ORIGIN

Query Match	89.6%	Score 22.4;	DB 14;	Length 178088;
Best Local Similarity	95.8%;	Pred. No. 3.8e+02;		
Matches 23;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY      2 AAAAAAAAAATCGCAACAACTCT 25
        |||||
Db     115802 AAAAAAAAAATCGCAAAAAAATCT 115825
```


chromosome 4p16.3
 DNA Ref. 5 (3), 177-186 (1998)
 PUBMED 9734812
 REFERENCE 2 (bases 1 to 7300)
 AUTHORS Hadano, S.
 TITLE Direct Submission
 JOURNAL Submitted (13-JAN-1997) Shinji Hadano, Japan Science and Technology Corporation, NeuroGenes Project, ICORP, Univ. of Tokai School of Med., Bohesida, Isehara, Kanagawa 259-1193, Japan
 (E-mail:shinji@nga.med.u-tokai.ac.jp, Tel:81-463-91-5095, Fax:81-463-91-4993)

FEATURES
 source
 1..7300
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4p16.3"
 /clone="RES4-23A"
 /tissue_type="Brain"
 /dev_stage="adult"
 /note="located close to the Huntington's disease gene"
 1..257
 /number=1
 258..397
 /number=2
 262..1947
 /codon_start=1
 /product="SH3 binding protein"
 /protein_id="BAA1919.1"
 /db_xref="GI:1843392"
 /translation="MAEEMHWVPMAKIGAONLITMPGVAKAGYLHKGGTOLIL
 KMPLEFVIHKRCVYFKSTASPOGAFSLGVRVRAAETTSNNVPRKIHIS
 DKHRWFSASASBEERKSMALIREIFHFHKDLPIDTSBSSDTSPFGAVERPV
 DISLPYTDNDYEHDDSDYLEPDSPEPGLDALMHPAYPPVPYPRKPAFS
 DMPRAHSTISKGPGLPPPPKGLPVGGLAABDSKRDPLCPRAEPCPVATPRK
 MDPPLSTMPYAPGLRKPCCFRESASPEEPTPGHACSTSSAIAIMATSRNCDKL
 KSFHLSRGPPTSEPPVPYANKPKFKLTAEDPPREAMPGLFVPPVAPRPALKLPY
 PEAMRAPVALPPPEKPOLPHLORSPIDQSPFSFEPKRPQSDKVDGSDSEYK
 VPLNSVYNTTESCEVERLFRATSPRSGPOLGICINSSITKSGKVLVWDETSNKV
 RNYRIEFKDSKRYLEGEVLFVSGSVMEHYTHVLPSHQSLLLRHPYGTGPR"
 398..500
 /number=3
 501..618
 /number=5
 619..689
 /number=6
 690..778
 /number=7
 779..847
 /number=8
 848..1496
 /number=9
 1497..1611
 /number=10
 1612..1667
 /number=11
 1668..1749
 /number=12
 1750..1809
 /number=13
 1810..7300
 /number=14
 2380..2385
 polyA_signal
 5134..5139
 polyA_signal
 6242..6247
 polyA_signal
 6266..6271

ORIGIN
 Query Match 88.0%; Score 22; DB 8; Length 7300;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAA 22
 Db 6764 AAAAAAAAAATCGCAACAAA 6743

RESULT 45
 HSL247F6
 LOCUS
 DEFINITION Human DNA sequence from clone LA04NC01-247F6 on chromosome 4,
 complete sequence.
 ACCESSION Z68279
 VERSION Z68279.1 GI:1130691
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 1 (bases 1 to 22970)
 Dodsworth, S.
 REFERENCE Direct Submission
 TITL Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep/LA04NC01-247F6
 is from a Los Alamos flow-sorted 4 library VECTOR: sCos-1.

FEATURES
 source
 1..22970
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="LA04NC01-247F6"
 /clone_id="LA04NC01"

ORIGIN
 Query Match 88.0%; Score 22; DB 8; Length 22970;
 Best Local Similarity 100.0%; Pred. No. 9.5e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAA 22
 Db 269 AAAAAAAAAATCGCAACAAA 290

RESULT 46
 HSJ323A24
 LOCUS
 DEFINITION Human DNA sequence from clone R3-323A24 on chromosome 4, complete

sequence.
 ACCESSION AL121750 AL158131 AL158132 AL158133
 VERSION AL121750.22 GI:50250927
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 149969)
 AUTHORS Bray-Allen, S.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUL-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT On Jul 12, 2004 this sequence version replaced gi:29788971.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP3-323A24 is from the library RPC1-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2.

 FEATURES
 source
 Location/Qualifiers
 1..149969
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="R2PD:RPC1P704A24323"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="RP3-323A24"
 /clone_11b="RPC1-3"

 ORIGIN
 Query Match 88.0%; Score 22; DB 8; Length 149969;
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATGCAACAA 22
 Db 58350 AAAAAAAAAATGCAACAA 58371

 RESULT 47
 LOCUS AP000780/c
 DEFINITION Homo sapiens chromosome 11 clone RP11-869J1 map 11q22, WORKING
 ACCESSION AP000780
 VERSION AP000780.2 GI:8131586
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 160111)
 AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Homo sapiens 160,111 genomic DNA of 11q22
 JOURNAL Published Only in DataBase (1999)
 REFERENCE 2 (bases 1 to 160111)
 AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (25-NOV-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail: hattori@gsc.riken.go.jp)
 URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-42-778-9923, Fax: 81-42-778-9924
 COMMENT On May 31, 2000 this sequence version replaced gi:697618.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: <http://hgp.gsc.riken.go.jp/>
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11.
 Center clone name: RP11-869J1
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator RT-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 134422 bases at least Q40
 Consensus quality: 146152 bases at least Q30
 Consensus quality: 153200 bases at least Q20
 Insert size: 156511; sum-of-coverage
 Quality coverage: 4.28x in Q20 bases; sum-of-coverage

 NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved
 1
 12950 12849 contig of 12849 bp in length
 23168 contig of 10219 bp in length
 23269 33786 contig of 10518 bp in length
 33887 43613 contig of 9727 bp in length
 43714 53115 contig of 9202 bp in length
 53216 60451 contig of 7236 bp in length
 60552 68431 contig of 7880 bp in length
 68532 75848 contig of 7317 bp in length
 75949 83172 contig of 7224 bp in length
 83273 88814 contig of 5542 bp in length
 88915 95974 contig of 7060 bp in length
 96075 100955 contig of 4881 bp in length
 101056 106548 contig of 4593 bp in length
 106649 110935 contig of 4287 bp in length
 111036 115006 contig of 3971 bp in length
 115107 117582 contig of 2476 bp in length
 117683 121247 contig of 3565 bp in length
 121348 124763 contig of 3416 bp in length
 124864 125267 contig of 404 bp in length
 125368 128895 contig of 3528 bp in length
 128996 132134 contig of 3139 bp in length
 134235 134561 contig of 2327 bp in length
 134662 136857 contig of 2196 bp in length
 136958 140689 contig of 3732 bp in length
 140790 141922 contig of 1133 bp in length
 142023 144760 contig of 2738 bp in length
 144861 145910 contig of 1050 bp in length

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 4, clone RP11-11012
Unpublished
2 (bases 1 to 196476)
Biren, B., Linton, J., Barna, N., Bastien, V., Bede, F.,
Anderson, S., Baldoni, J., Barna, N., Bastien, V., Bede, F.,
Boguslavsky, L., Boukhaltier, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karats, A.,
Klein, R., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheeters, R.,
McDermid, J., Menus, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 196476)
Biren, B., Linton, J., Barna, N., Bastien, V., Bede, F.,
Anderson, S., Baldoni, J., Barna, N., Bastien, V., Bede, F.,
Boguslavsky, L., Boukhaltier, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karats, A.,
Klein, R., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheeters, R.,
McDermid, J., Menus, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2000 this sequence version replaced gi:7582724.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 11.0.12
Center clone name: 11.0.12

Summary Statistics
Sequencing vector: M13, M7815, 100% of reads
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 173638 bases at least Q40
Consensus quality: 185088 bases at least Q30
Consensus quality: 189698 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 192576; sum-of-contrigs

Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 3.5 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently
consists of 40 contrigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contrigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1	1039:	contrig of 1039 bp	in length
1040	1139:	gap of 100 bp	
1140	2521:	contrig of 1382 bp	in length
2522	2621:	gap of 100 bp	
2622	3922:	contrig of 1301 bp	in length
3923	4022:	gap of 100 bp	
4023	5860:	contrig of 1838 bp	in length
5861	5960:	gap of 100 bp	
5961	7150:	contrig of 1190 bp	in length
7151	7250:	gap of 100 bp	
7251	8537:	contrig of 1287 bp	in length
8538	8637:	gap of 100 bp	
8638	10002:	contrig of 1355 bp	in length
10003	10102:	gap of 100 bp	
10103	11693:	contrig of 1591 bp	in length
11694	11793:	gap of 100 bp	
11794	13150:	contrig of 1357 bp	in length
13151	13250:	gap of 100 bp	
13251	14829:	contrig of 1679 bp	in length
14930	15029:	gap of 100 bp	
15030	17353:	contrig of 2324 bp	in length
17354	17453:	gap of 100 bp	
17454	19732:	contrig of 2279 bp	in length
19733	19832:	gap of 100 bp	
19833	21517:	contrig of 1785 bp	in length
21518	21717:	gap of 100 bp	
21718	24141:	contrig of 2424 bp	in length
24142	24241:	gap of 100 bp	
24242	26366:	contrig of 2125 bp	in length
26367	26466:	gap of 100 bp	
26467	29398:	contrig of 2932 bp	in length
29399	29498:	gap of 100 bp	
29499	32487:	contrig of 2969 bp	in length
32488	32587:	gap of 100 bp	
32588	35807:	contrig of 3220 bp	in length
35808	35907:	gap of 100 bp	
35908	38255:	contrig of 2348 bp	in length
38256	38355:	gap of 100 bp	
38356	42361:	contrig of 4006 bp	in length
42362	42461:	gap of 100 bp	
42462	45543:	contrig of 3082 bp	in length
45544	45643:	gap of 100 bp	
45644	49921:	contrig of 4278 bp	in length
49922	50021:	gap of 100 bp	
50022	53907:	contrig of 3866 bp	in length
53908	54007:	gap of 100 bp	
54008	58670:	contrig of 4663 bp	in length
58671	58770:	gap of 100 bp	
58771	63578:	contrig of 4808 bp	in length
63579	63678:	gap of 100 bp	
63679	67332:	contrig of 3654 bp	in length
67333	67432:	gap of 100 bp	
67433	72469:	contrig of 5037 bp	in length
72470	72569:	gap of 100 bp	
72570	76570:	contrig of 4101 bp	in length
76571	76770:	gap of 100 bp	
76771	82191:	contrig of 5421 bp	in length
82192	82291:	gap of 100 bp	
82292	88555:	contrig of 6264 bp	in length
88556	88655:	gap of 100 bp	
88656	96448:	contrig of 7793 bp	in length
96449	96548:	gap of 100 bp	


```

* 53740 58600: contig of 4861 bp in length
* 58601 58700: gap of unknown length
* 58701 63455: contig of 4755 bp in length
* 63456 63555: gap of unknown length
* 63556 70237: contig of 6682 bp in length
* 70238 70337: gap of unknown length
* 70338 75337: contig of 4900 bp in length
* 75338 80813: contig of 5476 bp in length
* 80814 80914: gap of unknown length
* 80914 85882: contig of 4969 bp in length
* 85883 85983: gap of unknown length
* 85983 91782: contig of 5800 bp in length
* 91783 91882: gap of unknown length
* 91883 100458: contig of 8576 bp in length
* 100459 100558: gap of unknown length
* 100559 107964: contig of 7406 bp in length
* 107965 108064: gap of unknown length
* 108065 116061: contig of 7997 bp in length
* 116062 116161: gap of unknown length
* 116162 125931: contig of 9770 bp in length
* 125932 126031: gap of unknown length
* 126032 137451: contig of 11420 bp in length
* 137452 137551: gap of unknown length
* 137552 154447: contig of 16886 bp in length
* 154448 154547: gap of unknown length
* 154548 174551: contig of 20004 bp in length
* 174552 174651: gap of unknown length
* 174652 199255: contig of 24604 bp in length.

```

FEATURES

```

source 1.199255 Location/Qualifiers
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="4"
        /clone="RP11-808B21"
misc_feature 1..1044
            /note="assembly_name:Contig11"
gap 1045..1144
            /estimated_length=unknown
misc_feature 1145..2445
            /note="assembly_name:Contig14"
gap 2446..2545
            /estimated_length=unknown
misc_feature 2546..4172
            /note="assembly_name:Contig15"
gap 4173..4272
            /estimated_length=unknown
misc_feature 4273..6230
            /note="assembly_name:Contig16"
gap 6231..6330
            /estimated_length=unknown
misc_feature 6331..8431
            /note="assembly_name:Contig17"
gap 8432..8531
            /estimated_length=unknown
misc_feature 8532..10262
            /note="assembly_name:Contig18"
gap 10263..10362
            /estimated_length=unknown
misc_feature 10363..11910
            /note="assembly_name:Contig19"
gap 11911..12010
            /estimated_length=unknown
misc_feature 12011..13067
            /note="assembly_name:Contig20"
gap 13068..13167
            /estimated_length=unknown
misc_feature 13168..14858
            /note="assembly_name:Contig21"
gap 14859..14958
            /estimated_length=unknown
misc_feature 14959..17103

```

```

gap /note="assembly_name:Contig22"
    17104..17203
    /estimated_length=unknown
misc_feature 17204..19088
    /note="assembly_name:Contig23"
    19089..19188
    /estimated_length=unknown
misc_feature 19189..22323
    /note="assembly_name:Contig24"
    22324..22423
    /estimated_length=unknown
misc_feature 22424..24304
    /note="assembly_name:Contig25"
    24305..24404
    /estimated_length=unknown
misc_feature 24405..27368
    /note="assembly_name:Contig26"
    27369..27468
    /estimated_length=unknown
misc_feature 27469..29295
    /note="assembly_name:Contig27"
    29296..29395
    /estimated_length=unknown
misc_feature 29396..32576
    /note="assembly_name:Contig28"
    32577..32676
    /estimated_length=unknown
misc_feature 32677..36828
    /note="assembly_name:Contig29"
    36829..36928
    /estimated_length=unknown
misc_feature 36929..41091
    /note="assembly_name:Contig30"
    41092..41191

```

```

Query Match      88.0%; Score 22; DB 14; Length 199255;
Best Local Similarity 100.0%; Pred.No. 5e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACCAA 22
Db 94338 AAAAAAAAAATCGCAACCAA 94359

```

```

RESULT 50
BV304791/c 460 bp DNA linear STS 26-JAN-2005
LOCUS BV304791.1
DEFINITION BV304791.1 TO Alaskan Malamute Canis familiaris STS genomic.
ACCESSION BV304791
VERSION BV304791.1 GI:57467026
KEYWORDS STS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
REFERENCE 1 (bases 1 to 460)
AUTHORS Lindblad-Toh K.
TITLE The genome sequence of Canis familiaris
JOURNAL Unpublished (2004)
COMMENT

```

Contact: Kerstin Lindblad-Toh
 Whitehead Institute for Biomedical Research, Center for Genome
 Research
 320 Charles Street, Cambridge, MA 02141, USA
 Tel: 6172521477
 Fax: 6172580903
 Email: kersli@genome.wi.mit.edu
 Primer A: No sequence submitted
 Primer B: No sequence submitted
 STS size: 460
 Protocol:

WGS-discovery (WGS) :

Paired-end low-coverage whole genome shotgun reads were generated from 9 breeds (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador Retriever, English Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese Water Dog -100,000 each) and five other canids (Chinese, Alaskan, Indian and Spanish Gray Wolf as well as the Californian Coyote).

The WGS reads were placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 863872 reads were annotated as STSs and 485941 SNPs were annotated with alleles from the boxer and the breed or canid from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

WGA-discovery (WGA) of Boxer/Poodle SNPs:

A second set of SNPs was generated using a similar methodology except that the contigs from the 1.5x poodle assembly (Kirkness 2003) were used instead of WGS reads. Since this sequence lacked base quality scores, arbitrary quality scores of phred 40 were assigned before the poodle sequence was placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated with alleles from the boxer and the poodle. The validation rate for these SNPs was estimated at approximately 100%.

Internal-WGA-discovery (I-WGA) :

A third set of SNPs were discovered by comparing reads in the WGA assembly. SNPs were defined as mismatch positions that had a base quality of >= 30 on both reads in a region that aligned without gaps, and with at most one additional mismatch in the ten flanking bases. For each allele, at least one additional read had to confirm it. 731476 SNPs were annotated with alleles between the two boxer alleles. The validation rate for these SNPs was estimated at approximately 100%.

FEATURES

source

```
1..460
    location/Qualifiers
        /organism="Canis familiaris"
        /mol_type="genomic DNA"
        /strain="AlaskanMalamute"
        /db_xref="taxon:9615"
        /map="1 22-372 20285083-20284733"
        /clone_lib="AlaskanMalamute"
        <1..>460
```

STS

ORIGIN

Query Match

Best Local Similarity 87.2%; Score 21.8; DB 10; Length 460;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25

DB 270 AAGAAAAAAATCCCAACAAATCT 246

Search completed: December 14, 2005, 11:10:02
Job time : 878.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 ; Search time 203.2 Seconds

(Without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-1

Perfect score: 25
Sequence: 1 aaaaaaaaaatcgcaacaactc 25

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 150 summaries

Database :
1: N_Geneseq_21.*
2: geneeqn1908.*
3: geneeqn1908.*
4: geneeqn2000.*
5: geneeqn2001.*
6: geneeqn2002.*
7: geneeqn2003.*
8: geneeqn2003.*
9: geneeqn2003.*
10: geneeqn2003.*
11: geneeqn2003.*
12: geneeqn2004.*
13: geneeqn2004.*
14: geneeqn2005.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.4	93.6	142318	11	ACN44850 Human gen
2	22	88.0	4068	4	AA186707 Human pol
3	22	88.0	7300	14	AD280551 SH3-domain
4	22	88.0	13388	10	ADC71368 Human col
5	22	88.0	22970	10	ADK70082 Mutant hu
6	22	88.0	22970	10	ADK70081 Wild type
7	21.8	87.2	6162	6	ABL92315 Chemical
8	21.8	87.2	6226	6	ABK43074 Genomic s
9	21.8	87.2	6226	9	AD61230 Connectiv
10	21.8	87.2	7145	6	ABL92234 Chemical
11	21.8	87.2	7145	6	AA022321 Chemical
12	21.8	87.2	7928	6	ABL22095 Human imm
13	21.8	87.2	7928	6	ABK31169 Signal tr
14	21.8	87.2	7928	6	ABL70124 Chemical
15	21.8	87.2	7928	6	AA61056 Human gen
16	21.8	87.2	31024	10	ABV75372 Human IGF
17	21.8	87.2	37736	4	AAK6139 Human imm
18	21.8	87.2	180227	13	ABD33268 Human can
19	21.4	85.6	1710	6	ABL34385 Human imm

20	21	84.0	230	7	AD531291
21	21	84.0	230	7	ADY36679
22	20.8	83.2	544	6	ABQ25827 Oligonuc
23	20.8	83.2	544	6	ABQ25826 Oligonuc
24	20.8	83.2	655	4	AAK56935 Human imm
25	20.8	83.2	722	6	ABQ35161 Oligonuc
26	20.8	83.2	722	6	ABQ35160 Oligonuc
27	20.8	83.2	730	6	ABQ15220 Oligonuc
28	20.8	83.2	730	6	ABQ15221 Oligonuc
29	20.8	83.2	1096	4	AAK69138 Human imm
30	20.8	83.2	1834	6	AA515007 DNA encod
31	20.8	83.2	2000	11	ACL38225
32	20.8	83.2	2000	11	ACL38226
33	20.8	83.2	2000	12	ADJ40625
34	20.8	83.2	5763	6	ABL32182 Human imm
35	20.8	83.2	6127	6	ABL33615 Human imm
36	20.8	83.2	6221	4	AA546503 Tumour su
37	20.8	83.2	9851	6	ABL32795
38	20.8	83.2	9859	13	AD589440
39	20.8	83.2	9859	13	AD589714
40	20.8	83.2	10144	6	ABL49393
41	20.8	83.2	13038	6	ABL33275 Human imm
42	20.8	83.2	14542	6	ABK31234 Signal tr
43	20.8	83.2	14542	6	ABK70191 Chemical
44	20.8	83.2	14542	6	AA61147 Human gen
45	20.8	83.2	15872	4	AA646519 Tumour su
46	20.8	83.2	17968	14	ADM10346
47	20.8	83.2	17968	14	ADM10488
48	20.8	83.2	32132	4	AA135943 Human mns
49	20.8	83.2	32132	4	AA107060 Human rep
50	20.8	83.2	32132	8	ABK58931 CDNA enco
51	20.8	83.2	32132	12	ADJ29681 Human mns
52	20.8	83.2	32169	4	ABK43076 Genomic s
53	20.8	83.2	32169	9	AD61232 Connectiv
54	20.8	83.2	58181	8	AB274619 Secreted
55	20.8	83.2	58181	10	ADC21010 Human sec
56	20.8	83.2	58181	10	AB268140 Human sec
57	20.8	83.2	110000	10	ABE11169_0
58	20.4	81.6	376	5	ABV61205 Human pro
59	20.4	81.6	585	6	ABQ19872 Oligonuc
60	20.4	81.6	585	6	ABQ19873 Oligonuc
61	20.4	81.6	610	5	ABV33963 Human pro
62	20.4	81.6	850	6	ABQ48654 Oligonuc
63	20.4	81.6	850	6	ABQ48655 Oligonuc
64	20.4	81.6	904	6	ABQ38761 Oligonuc
65	20.4	81.6	904	6	ABQ38760 Oligonuc
66	20.4	81.6	977	13	ADR37702 African g
67	20.4	81.6	1112	14	ABE28608 Maize poi
68	20.4	81.6	2607	5	AAE97848 Human neu
69	20.4	81.6	3140	4	ABL27038 Drosophi
70	20.4	81.6	6587	6	ABL32385 Human imm
71	20.4	81.6	6700	6	ABN80304 Human che
72	20.4	81.6	8420	4	AA546658 Tumour su
73	20.4	81.6	8420	6	ABK31325 Signal tr
74	20.4	81.6	11049	6	ABV92220 Chemical
75	20.4	81.6	11049	6	ABL49323 Human pol
76	20.4	81.6	20345	10	ADC86252 Human GPC
77	20.4	81.6	26390	4	AAK65971 Human imm
78	20.4	81.6	30821	11	ACN44698 Human gen
79	20.4	81.6	86574	6	ABK83560 Human cdn
80	20.4	81.6	86574	13	ADR52822 Drug ther
81	20.4	81.6	147300	12	ADP45593 Human Rho
82	20.4	81.6	147700	14	ADX98570 Human gna
83	20.2	80.8	329	5	ABA11022 Human ner
84	20.2	80.8	372	6	ABK78971 Bacillus
85	20.2	80.8	389	5	ABA14388 Human ner
86	20.2	80.8	399	14	ADW83459 MAP3K9 ma
87	20.2	80.8	554	6	ABQ42893 Oligonuc
88	20.2	80.8	554	6	ABQ42892 Oligonuc
89	20.2	80.8	684	6	ABQ44984 Oligonuc
90	20.2	80.8	684	6	ABQ44985 Oligonuc
91	20.2	80.8	806	6	ABQ17664 Oligonuc
92	20.2	80.8	806	6	ABQ17665 Oligonuc

AD631291	Human gen
ADY36679	H1RA geno
ABQ25827	Oligonuc
ABQ25826	Oligonuc
AAK56935	Human imm
ABQ35161	Oligonuc
ABQ35160	Oligonuc
ABQ15220	Oligonuc
ABQ15221	Oligonuc
AAK69138	Human imm
AA515007	DNA encod
ACL38225	Rice stre
ACL38226	Rice stre
ADJ40625	Plant cDN
ABL32182	Human imm
ABL33615	Human imm
AA546503	Tumour su
ABL32795	Human imm
AD589440	Oligonuc
AD589714	Oligonuc
ABL49393	Human pol
ABL33275	Human imm
ABK31234	Signal tr
ABK70191	Chemical
AA61147	Human gen
AA646519	Tumour su
ADM10346	Colon pro
ADM10488	Colon pro
AA135943	Human mns
AA107060	Human rep
ABK58931	CDNA enco
ADJ29681	Human mns
ABK43076	Genomic s
AD61232	Connectiv
AB274619	Secreted
ADC21010	Human sec
AB268140	Human sec
ABE11169	Human tra
ABV61205	Human pro
ABQ19872	Oligonuc
ABQ19873	Oligonuc
ABV33963	Human pro
ABQ48654	Oligonuc
ABQ48655	Oligonuc
ABQ38761	Oligonuc
ABQ38760	Oligonuc
ADR37702	African g
ABE28608	Maize poi
AAE97848	Human neu
ABL27038	Drosophi
ABL32385	Human imm
ABN80304	Human che
AA546658	Tumour su
ABK31325	Signal tr
ABV92220	Chemical
ABL49323	Human pol
ADC86252	Human GPC
AAK65971	Human imm
ACN44698	Human gen
ABK83560	Human cdn
ADR52822	Drug ther
ADP45593	Human Rho
ADX98570	Human gna
ABA11022	Human ner
ABK78971	Bacillus
ABA14388	Human ner
ADW83459	MAP3K9 ma
ABQ42893	Oligonuc
ABQ42892	Oligonuc
ABQ44984	Oligonuc
ABQ44985	Oligonuc
ABQ17664	Oligonuc
ABQ17665	Oligonuc

93	20.2	80.8	857	6	ABQ34087	Abq34087 Oligonuc1
C 94	20.2	80.8	857	6	ABQ34086	Abq34086 Oligonuc1
C 95	20.2	80.8	959	6	ABQ26660	Abq26660 Oligonuc1
C 96	20.2	80.8	959	6	ABQ26661	Abq26661 Oligonuc1
C 97	20.2	80.8	973	6	ABQ45761	Abq45761 Oligonuc1
C 98	20.2	80.8	973	6	ABQ45760	Abq45760 Oligonuc1
C 99	20.2	80.8	978	6	ABQ40294	Abq40294 Oligonuc1
C 100	20.2	80.8	978	6	ABQ40295	Abq40295 Oligonuc1
C 101	20.2	80.8	1292	4	ABN06371	Abn06371 Human cDN
C 102	20.2	80.8	1292	4	ABV83708	Abv83708 Human pol
C 103	20.2	80.8	1390	5	ABA21084	Ab21084 Human ner
C 104	20.2	80.8	1390	5	ABA21083	Ab21083 Human ner
C 105	20.2	80.8	1505	6	ABL40965	Ab140965 Polypepti
C 106	20.2	80.8	1732	4	ABL15627	Ab115627 Drosophil
C 107	20.2	80.8	2108	6	ABN03915	Abn03915 Human con
C 108	20.2	80.8	2221	6	ABL70431	Ab170431 Chemical1
C 109	20.2	80.8	2221	6	AAE61374	AAe61374 Human gen
C 110	20.2	80.8	3277	10	ADE57827	Ade57827 Rat gene
C 111	20.2	80.8	3605	4	AAH29886	Aah29886 C albican
C 112	20.2	80.8	4183	4	ABU15626	Ab115626 Drosophil
C 113	20.2	80.8	5152	6	ABU49348	Ab149348 Human pol
C 114	20.2	80.8	5108	6	ABU32636	Ab132636 Human imm
C 115	20.2	80.8	5312	6	ABU32197	Ab132197 Human imm
C 116	20.2	80.8	5714	6	ABRK3196	Ab131196 Signal tr
C 117	20.2	80.8	5714	6	ABU70163	Ab170163 Chemical1
C 118	20.2	80.8	5714	6	AAE61108	AAe61108 Human gen
C 119	20.2	80.8	5928	6	ABU34087	Ab134087 Human imm
C 120	20.2	80.8	6326	6	ABU92287	Ab192287 Chemical1
C 121	20.2	80.8	6326	6	ABU49358	Ab149358 Human pol
C 122	20.2	80.8	6326	6	AAE22338	AAe22338 Chemical1
C 123	20.2	80.8	7406	6	ABN80157	Abn80157 Human che
C 124	20.2	80.8	7680	3	AAE89439	AAe89439 14-3-3 si
C 125	20.2	80.8	8700	6	ABU33003	Ab133003 Human imm
C 126	20.2	80.8	10026	6	ABU33876	Ab133876 Human imm
C 127	20.2	80.8	10034	6	ABU06535	Ab06535 Human 14-
C 128	20.2	80.8	10034	11	ADM83712	Adm83712 Human cyc
C 129	20.2	80.8	10872	6	ABK31388	Ab131388 Signal tr
C 130	20.2	80.8	10872	6	ABU70335	Ab170335 Chemical1
C 131	20.2	80.8	10872	6	AAE61288	AAe61288 Human gen
C 132	20.2	80.8	12138	6	ABK40033	Abk40033 Human che
C 133	20.2	80.8	12138	6	ABU34328	Ab134328 Human imm
C 134	20.2	80.8	12337	6	ABU34359	Ab134359 Human imm
C 135	20.2	80.8	12423	10	ADBS4081	Adbs4081 Pretreatm
C 136	20.2	80.8	12423	13	ADBS89235	Adbs89235 Oligonuc1
C 137	20.2	80.8	13732	6	ABU33820	Ab133820 Human imm
C 138	20.2	80.8	15416	6	ABU34230	Ab134230 Human imm
C 139	20.2	80.8	15416	6	ABK31526	Abk31526 Signal tr
C 140	20.2	80.8	15416	6	ABU70497	Ab170497 Chemical1
C 141	20.2	80.8	15416	6	AAE61452	AAe61452 Human gen
C 142	20.2	80.8	15698	6	ABU34140	Ab134140 Human imm
C 143	20.2	80.8	15825	6	ABU33242	Ab133242 Human imm
C 144	20.2	80.8	15825	6	ABU34558	Ab134558 Human met
C 145	20.2	80.8	15825	7	ADBS99819	Adbs99819 Bisluplit
C 146	20.2	80.8	16439	6	ABU32886	Ab132886 Human imm
C 147	20.2	80.8	17203	13	ADBS89745	Adbs89745 Oligonuc1
C 148	20.2	80.8	17970	6	ABO67034	Ab067034 Human ang
C 149	20.2	80.8	18605	6	ABN84505	Abn84505 Human smo
C 150	20.2	80.8	18605	6	ACA60862	ACA60862 Human smo

ALIGNMENTS

RESULT 1
ID ACN44850/c
AC ACN44850; standard; DNA; 142318 BP.
XX
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG33122.
XX

KW	Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO2003073826-A2.
XX	
PD	12-SEP-2003.
XX	
PF	28-FEB-2003; 2003WO-US006235.
XX	
PR	01-MAR-2002; 2002US-00087192.
XX	
PA	(SAGR-) SAGRES DISCOVERY.
XX	
PI	Morris DW;
XX	
DR	WPI; 2003-328604/31.
XX	
PT	Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT	comprises a nucleotide sequence.
PS	Claim 1; SEQ ID NO 1504; OPD; English.
XX	
CC	The present invention relates to novel DNA and protein sequences which
CC	are associated with carcinomas. The sequences are useful for: (i) for
CC	screening drug candidates; (ii) for screening of bioactive agent capable
CC	of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC	a bioactive agent capable of modulating the activity of CAP; (iv) for
CC	evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC	carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC	carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC	(x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC	determining Carcinoma Associated (CA) gene copy number. In addition, the
CC	CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC	carcinoma including lymphoma. The present sequence is one such CA coding
CC	sequence. Note: This patent is an equivalent to basic patent
CC	US2002182586A1, for which no sequence data was published
XX	
SEQ	Sequence 142318 BP; 38833 A; 25825 C; 26759 G; 47076 T; 0 U; 3825 Other;
XX	
Query Match	93.6%; Score 23.4; DB 1; Length 142318;
Best Local Similarity	96.0%; Pred. No. 1.2e+02;
Matches	24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY	1 AAAAAAAAAATCGCAACAAATCT 25
DB	121397 AAAAAAAAAATCGCAACAAATCT 121373
RESULT 2	
ID	AA186707 standard; cDNA; 4068 BP.
AC	AA186707;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 6767.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukemia;
KW	nervous system disorders; arthritis; inflammation; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200164835-A2.
XX	
PD	07-SEP-2001.
XX	
PF	26-FEB-2001; 2001WO-US004927.
XX	
PR	28-FEB-2000; 2000US-00515126.

```

PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AAO06776.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 1; SEQ ID NO 6767; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukemia, nervous system disorders, arthritis and
XX inflammation. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4068 BP; 1134 A; 1070 C; 1053 G; 811 T; 0 U; 0 Other;
XX
XX Query Match 88.0%; Score 22; DB 4; Length 4068;
XX Best Local Similarity 100.0%; Pred. No. 3.5e+02;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
OY 1 AAAAAAAAAATCGCAACAA 22
XX ||||||||||||||||
XX Db 536 AAAAAAAAAATCGCAACAA 557
XX
RESULT 3
AD280551/C
ID AD280551 standard; cDNA; 7300 BP.
XX
XX AD280551;
XX
XX 14-JUL-2005 (first entry)
XX
XX SH3-domain binding protein 2 (SH3BP2) cDNA.
XX
XX DE
XX DE cardiant; vasotropic; gene therapy; diagnosis; prognosis;
XX KW gene expression; coronary artery disease; cardiant; vasotropic;
XX KW cardiovascular disease; ds; SH3BP2.
XX
XX Homo sapiens.
XX
XX WO2005040422-A2.
XX
XX 06-MAY-2005.
XX
XX 15-OCT-2004; 2004WO-EP011651.
XX
XX 16-OCT-2003; 2003US-0511784P.
XX PR 27-MAY-2004; 2004US-0574818P.
XX
XX (NOVS ) NOVARTIS AG.
XX PA (NOVS ) NOVARTIS PHARMA GMBH.
XX
XX Chibout S, Gras P, Vonderscher J;
XX
XX WPI; 2005-355863/36.
XX
XX Identifying or predicting (the predisposition of) CAD, monitoring a
XX subject identified as having CAD before and after treatment or the

```

```

Pr progression or severity of CAD by determining specific peptide/gene
Pt expression levels.
Ps Claim 7; SEQ ID NO 17; 136pp; English.
Pp
Pq The invention describes a method of identifying or predicting (the
Pr predileposition of) coronary artery disease (CAD), monitoring a subject
Ps identified as having CAD before and after treatment or the progression or
Pp severity of CAD comprises determining the level of one or more peptides
Pr or gene expression of at least one gene given in the specification in a
Pp subject to provide a first value and in a control or reference standard
Pr to provide a second value, and comparing whether there is a difference
Pp between the first value and second value. Also described are: screening
Pr candidate agents for use in treatment of CAD; treating or preventing CAD;
Pp manufacture of a medicament for the treatment or prevention of CAD
Pr comprising an agent that can induce a decrease in the level of gene
Pp expression, synthesis, or activity of at least one gene or gene
Pr expression products; and a kit for the identifying or predicting the
Pp predileposition CAD in a subject comprising: instructions for determining
Pr the peptide level and/or level of gene expression; control or reference
Pp standard peptide level and/or level of gene expression from a normal
Pr subject(s) without CAD for at least one gene or peptide cited above. A
Pp substance comprising an agent that can induce a decrease in the level of
Pr gene expression, synthesis, or activity of at least one gene or gene
Pp expression product or the level of at least one disease over control
Pr and/or predominant in Disease peptide and/or induce an increase in the
Pp level of at least one control over Disease and/or Predominant in Control
Pr peptide and/or a decrease in gene expression, synthesis, or activity of
Pp at least one gene or gene expression product is useful for manufacturing
Pr a medicament for the treatment or prevention of CAD. The method is useful
Pp for identifying or predicting (the predileposition of) CAD, monitoring a
Pr subject identified as having CAD before and after treatment or monitoring
Pp the progression or severity of CAD, thus treating CAD. This sequence
Pr represents a PTEN induced putative kinase 1 polynucleotide, a highly
Pp predictive gene used to determine the severity of coronary artery
Pr disease.
Pq
Pp Sequence 7300 BP; 1478 A; 2074 C; 1983 G; 1765 T; 0 U; 0 Other;
Pq
Pp
Pq Query Match 88.0%; Score 22; DB 14; Length 7300;
Pp Best Local Similarity 100.0%; Pred. NO. 3.5e+02;
Pp Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Pp
Pq Oy 1 AAAAAAAAAATCGCAACAAA 22
Pp ||||||||||||||||||||
Pq Db 6764 AAAAAAAAAATCGCAACAAA 6743
Pp
Pp RESULT 4
Pp ADC71368/C
Pp ID ADC71368 standard; cDNA; 13388 BP.
Pp
Pp AC ADC71368;
Pp XX
Pp XX 18-DEC-2003 (first entry)
Pp XX
Pp DE Human colon specific cDNA sequence DEX0235_64 (SegID 64).
Pp
Pp KW human; gene; sg; neoplastic colorectal; colon cancer;
Pp KW non-cancerous disease; gene therapy; transgenic; DEX0235_64.
Pp
Pp OS Homo sapiens.
Pp
Pp PM WO2003020934-A1.
Pp
Pp PD 13-MAR-2003.
Pp
Pp PF 29-AUG-2002; 2002WO-US027737.
Pp
Pp PR 31-AUG-2001; 2001US-0316258P.
Pp
Pp PA (DIAD-) DIADEXUS INC.
Pp
Pp

```

PI Sun Y, Liu C, Ghosh KG;
 XX WPI, 2003-300891/29.
 DR P-PSDB; ADCT1301.
 XX Novel colon specific polypeptides and nucleic acids, useful for
 PT identifying, diagnosing, monitoring, staging, imaging and treating colon
 PT cancer and non-cancerous disease states in colon tissue.
 XX
 PS Claim 1; SEQ ID NO 64; 262bp; English.
 XX
 CC This invention relates to novel nucleic acid molecules and the encoded
 CC polypeptides, which are present in normal and neoplastic colorectal
 CC cells. Specifically, it refers to antibodies of these colon specific
 CC polypeptides, as well as antagonists and agonists thereof that can be
 CC used to treat colon cancer and also non-cancerous diseases states of the
 CC colon. The present invention describes methods useful for the diagnosis
 CC and monitoring of colon cancer metastases in a patient, by determining
 CC the concentration of these colon specific proteins in a patient sample.
 CC Furthermore, they are also used for gene therapy purposes, the production
 CC of transgenic animals and cells, as well as producing engineered colon
 CC tissue for treatment and research. This polynucleotide sequence is a
 CC human colon specific cDNA sequence of the invention.
 CC
 SQ Sequence 13388 BP; 3129 A; 3503 C; 3340 G; 3414 T; 0 U; 2 Other;
 XX
 Query Match 88.0%; Score 22; DB 10; Length 13388;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAAA 22
 DB 6765 AAAAAAAAAATCGCAACAAA 6744
 XX
 RESULT 5
 ADK70082
 ID ADK70082 standard; DNA; 22970 BP.
 XX
 AC ADK70082;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Mutant human SH3 binding protein 2 genomic sequence.
 XX
 KM ds; gene; osteopathic; cytostatic; gene therapy; SH3-binding protein;
 KM SH3BP2; mutation; diagnosis; bone homeostasis; cherubism; bone tumor.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2003025197-A2.
 XX
 PD 27-MAR-2003.
 XX
 PF 01-FEB-2002; 2002WO-US019164.
 XX
 PR 02-FEB-2001; 2001US-0266129P.
 XX
 PA (HARD) HARVARD COLLEGE.
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
 XX
 PI Tiziani V, Reichenberger E, Ueki Y, Olsen BR;
 XX WPI, 2003-371820/35.
 DR
 XX New mutant SH3BP2 nucleic acid molecule or polypeptide, useful for
 PT diagnosing and treating disorders of bone homeostasis, such as cherubism
 PT or bone tumor.
 XX
 PS Claim 1; SEQ ID NO 26; 70pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule comprising a

CC mutant SH3-binding protein and its encoding DNA, nucleic acid molecules
 CC which hybridize under stringent conditions to it, a nucleic acid
 CC molecules that differ from it in codon sequence due to the degeneracy of
 CC the genetic code; and complements of these, provided that the nucleic
 CC acid molecule is not a human 'WT' SH3BP2 full length sequence. The mutant
 CC is especially selected from a genetic mutant domain; 'H' family mutation;
 CC 'K' family mutation; 'A,B' family mutation; 'C, F, U, M, O' family
 CC mutation; 'L' family mutation; 'G' family mutation; 'N' family mutation
 CC or genomic mutant SH3BP2; any of the mutant SH3BP2 exon 9 sequences and
 CC a genomic mutant SH3BP2 nucleic acid molecule. The nucleic acid molecule
 CC and polypeptide are useful for diagnosing and treating disorders of bone
 CC homeostasis, such as cherubism or a bone tumor. This sequence corresponds
 CC to the mutant genomic SH3-binding protein DNA sequence.
 CC
 SQ Sequence 22970 BP; 5407 A; 6744 C; 6741 G; 4069 T; 0 U; 9 Other;
 XX
 Query Match 88.0%; Score 22; DB 10; Length 22970;
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAAA 22
 DB 269 AAAAAAAAAATCGCAACAAA 290
 XX
 RESULT 6
 ADK70081
 ID ADK70081 standard; DNA; 22970 BP.
 XX
 AC ADK70081;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Wild type human SH3 binding protein 2 genomic sequence.
 XX
 KM ds; gene; osteopathic; cytostatic; gene therapy; SH3-binding protein;
 KM SH3BP2; mutation; diagnosis; bone homeostasis; cherubism; bone tumor.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2003025197-A2.
 XX
 PD 27-MAR-2003.
 XX
 PF 01-FEB-2002; 2002WO-US019164.
 XX
 PR 02-FEB-2001; 2001US-0266129P.
 XX
 PA (HARD) HARVARD COLLEGE.
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
 XX
 PI Tiziani V, Reichenberger E, Ueki Y, Olsen BR;
 XX WPI, 2003-371820/35.
 DR
 XX New mutant SH3BP2 nucleic acid molecule or polypeptide, useful for
 PT diagnosing and treating disorders of bone homeostasis, such as cherubism
 PT or bone tumor.
 XX
 PS Disclosure; SEQ ID NO 25; 70pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule comprising a
 CC mutant SH3-binding protein and its encoding DNA, nucleic acid molecules
 CC which hybridize under stringent conditions to it, a nucleic acid
 CC molecules that differ from it in codon sequence due to the degeneracy of
 CC the genetic code; and complements of these, provided that the nucleic
 CC acid molecule is not a human 'WT' SH3BP2 full length sequence. The mutant
 CC is especially selected from a genetic mutant domain; 'H' family mutation;
 CC 'K' family mutation; 'A,B' family mutation; 'C, F, U, M, O' family
 CC mutation; 'L' family mutation; 'G' family mutation; 'N' family mutation
 CC or genomic mutant SH3BP2; any of the mutant SH3BP2 exon 9 sequences and
 CC a genomic mutant SH3BP2 nucleic acid molecule. The nucleic acid molecule
 CC and polypeptide are useful for diagnosing and treating disorders of bone

CC homeostasis, such as chemubism or a bone tumor. This sequence corresponds
CC to the wild type genomic SH3-binding protein DNA sequence.
XX
SQ Sequence 22970 BP; 5407 A; 6748 C; 6745 G; 4070 T; 0 U; 0 Other;
Query Match 88.0%; Score 22; DB 10; Length 22970;
Best Local Similarity 100.0%; Pred.No.3.4e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACAAA 22
Db 269 AAAAAAAAAATCGCAACAAA 290
RESULT 7
ABL92315/C
ID ABL92315 standard; DNA; 6162 BP.
XX
AC ABL92315;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated DNA repair gene fragment complementary to#62.
XX
KM DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L3; PMS2;
KM L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4; DDT1L; FANCB;
KM XRCC8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;
KM Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
KM trichthodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.
XX
OS Unidentified.
XX
PN WO200181622-A2.
XX
PD 01-NOV-2001.
XX
PF 06-APR-2001; 2001WO-EP003972.
XX
PR 06-APR-2000; 2000DE-01019058.
XX
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01033529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-034446/04.
XX
PT New nucleic acid derived from genes associated with DNA repair, useful
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT cytosine methylation.
XX
PS Claim 1; SEQ ID NO 124; 25pp + Sequence Listing; English.
XX
CC The invention relates to nucleic acids containing a sequence of at least
CC 18 nucleotides of chemically treated DNA of genes associated with DNA
CC repair, and their complements. The invention also relates to nucleic
CC acids comprising at least 18 base pairs of the chemically pretreated DNA
CC of genes associated with DNA repair selected from PMS2L1, PMS2L2,
CC PMS2L3, PMS2L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,
CC RFC4, DDT1L, FANCB, or XRCC8. Nucleic acids of the invention and related
CC oligomers, are useful for diagnosis of diseases associated with gene
CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
CC immunodeficiency, trichthodystrophy, Fanconi's anaemia, solid tumours
CC and cancer, particularly by determining status of cytosine methylation
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records ABL92192-ABL92335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: The sequence data for this patent is
CC not represented in the specification, but is based on sequence

CC information supplied by the European Patent Office
XX
SQ Sequence 6162 BP; 1341 A; 178 C; 1476 G; 3167 T; 0 U; 0 Other;
Query Match 87.2%; Score 21.8; DB 6; Length 6162;
Best Local Similarity 92.0%; Pred.No.4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 3579 AAAAAAAAAATCGCAACAAATCT 3555
RESULT 8
ABK43074/C
ID ABK43074 standard; DNA; 6226 BP.
XX
AC ABK43074;
XX
DT 21-MAY-2002 (first entry)
XX
DE Genomic sequence #973 encoding novel human connective tissue polypeptide.
XX
KM Human; connective tissue related disorder; cancer; gene therapy;
KM cytosinatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200155343-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001322.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226797P.
PR 22-AUG-2000; 2000US-022681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 06-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 08-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 12-SEP-2000; 2000US-0232081P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239393P.
 PR 13-OCT-2000; 2000US-0239393P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.

PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCL INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI, 2001-565190/63.
 XX
 XX Nucleic acid encoding novel connective tissue associated polypeptides,
 PT used in diagnosing, preventing, treating or ameliorating a disorder such
 PT as cancer or rheumatoid arthritis.
 XX
 PS Disclosure; SEQ ID NO 1961; 673pp; English.
 XX
 CC The present invention relates to the isolation of novel human connective
 CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
 CC (DNA and genomic) sequences encoding them. The sequences of the
 CC invention are useful in the diagnosis, treatment, prevention and/or
 CC prognosis of diseases associated with connective tissue(s), including
 CC cancer. The polynucleotide sequences of the invention are also useful in
 CC gene therapy. ABK42102-ABK43116 represent genomic sequences encoding the
 CC novel human connective tissue related polypeptides. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6226 BP; 1500 A; 1552 C; 1518 G; 1656 T; 0 U; 0 Other;
 Query Match 87.2%; Score 21.8; DB 4; Length 6226;
 Best Local Similarity 92.0%; Pred. No. 4e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAATCGCAACAATCT 25
 Db 660 AAAAAAAAAATCGCAACAATCT 636
 RESULT 9
 ADB61230/c
 ID ADB61230 standard; DNA; 6226 BP.
 AC ADB61230;
 XX
 XX 04-DEC-2003 (first entry)
 DT
 XX
 DE Connective tissue related genomic DNA #973.
 XX
 XX cytosolic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;
 KW antiarteriosclerotic; immunosuppressive; antineumatic; antiarthritic;
 KW antiinflammatory; antiallergic; antiashtmatic; dermatological;

KM nephrotropic; virucide; fungicide; antibacterial; antiparasitic;
KM gene therapy; ds; connective tissues disorder; rheumatoid arthritis;
KM systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;
KM cancer metastasis; neoplasia; leukaemia; neurodegenerative disorder;
KM Alzheimer's disease; Parkinson's disease; cardiovascular disease;
KM atherosclerosis; myocarditis; cardiopulmonary bypass complication;
KM autoimmune disease; multiple sclerosis; allergic reaction; asthma;
KM rhinitis; eczema; inflammatory condition; Crohn's disease; nephritis;
KM gastrointestinal disorder; inflammatory bowel disease;
KM organ transplant rejection; immune system disorder; Bruton's disease;
KM X-linked lymphoproliferative syndrome;
KM B-cell lymphoproliferative disorder; HIV; AIDS; infection;
KM chromosome identification; chromosome mapping;
KM connective tissue related polynucleotide; gene; ds.
KM
XX Homo sapiens.
OS
XX
XX US2003054375-A1.
PN
XX
XX 20-MAR-2003.
PD
XX
XX
XX 07-MAR-2002; 2002US-00092154.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231422P.
PR 08-SEP-2000; 2000US-0231423P.
PR 08-SEP-2000; 2000US-0231433P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0233397P.
PR 14-SEP-2000; 2000US-0233398P.
PR 14-SEP-2000; 2000US-0233399P.
PR 14-SEP-2000; 2000US-0233400P.
PR 14-SEP-2000; 2000US-0233401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.

PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764847.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Barash SC,
XX
XX MPI: 2003-634869/60.
XX P-PSDB; ADB59748.
XX
XX
XX New connective tissue-related polypeptides and polynucleotides, useful
XX for treating, preventing and/or prognosing e.g. disorders of connective
XX tissue, (e.g. rheumatoid arthritis), cancers, cancer metastases and/or
XX neoplasias.
XX
XX
XX Disclosure; SEQ ID NO 1961; 248pp; English.
XX
XX
XX The invention describes an isolated nucleic acid molecule (1), which
XX comprises a sequence that is at least 95 % identical to a connective
XX tissue-related polynucleotide encoding connective tissue antigens (CTA).
XX
XX The polypeptide or polynucleotide is useful for preventing, treating, or
XX ameliorating medical conditions in a mammal. The connective tissue
XX polypeptides, polynucleotides and antibodies are particularly useful for
XX treating, preventing and/or prognosing disorders of connective tissues
XX (e.g. rheumatoid arthritis, discoid and systemic lupus erythematosus,
XX scleroderma, or Sjogren's syndrome), cancers, cancer metastases and/or
XX neoplasias (e.g. leukaemia), neurodegenerative disorders (e.g.
XX Alzheimer's disease, or Parkinson's disease), cardiovascular diseases
XX (e.g. atherosclerosis, myocarditis or cardiopulmonary bypass
XX complications), autoimmune diseases (e.g. systemic lupus erythematosus,
XX rheumatoid arthritis, or multiple sclerosis), allergic reactions (e.g.
XX

Query Match 87.2%; Score 21.8; DB 9; Length 6226;
Best Local Similarity 92.0%; Pred. No. 4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCGCAACAATCT 25
DB 660 AAAAAAAAAATCGCAACAATCT 636

RESULT 10
ABL92234/C
ID ABL92234 standard; DNA; 7145 BP.
XX
XX ABL92234;
XX
XX 01-JUN-2002 (first entry)
XX
XX
XX Chemically treated DNA repair gene fragment#22.
XX
XX DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3; PMS2;
XX L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4; DIT1L; FANCB;
XX XRCB8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;
XX Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
XX trichothiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.
XX
XX
XX Undifferentiated.
XX
XX
XX WO200181622-A2.
XX

PD 01-NOV-2001.
XX
XX
XX 06-APR-2001; 2001WO-EP003972.
XX
XX
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX
XX Olek A, Piepenbrock C, Berlin K,
XX
XX MPI: 2002-034446/04.
XX
XX
XX New nucleic acid derived from genes associated with DNA repair, useful
XX for diagnosis, e.g. of ataxia telangiectasia, by determination of
XX cytosine methylation.
XX
XX
XX Claim 1; SEQ ID NO 43; 25pp + Sequence Listing; English.
XX
XX
XX The invention relates to nucleic acids containing a sequence of at least
XX 18 nucleotides of chemically treated DNA of genes associated with DNA
XX repair, and their complements. The invention also relates to nucleic
XX acids comprising at least 18 base pairs of the chemically pretreated DNA
XX of genes associated with DNA repair selected from PMS2L1, PMS2L12,
XX PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, NUDT1, TDG, INPPL1,
XX RFC4, DIT1L, FANCB, XRCB8. Nucleic acids of the invention and related
XX oligomers, are useful for diagnosis of diseases associated with gene
XX repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
XX Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
XX immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours
XX and cancer, particularly by determining status of cytosine methylation
XX and/or by detecting single-nucleotide polymorphisms. Determination of
XX individual methylation patterns may allow development of individualised
XX therapies. The sequences given in records ABL92192-ABL92335 represent
XX chemically pre-treated DNA fragments from genes associated with DNA
XX repair, and their complements. Note: The sequence data for this patent is
XX not represented in the specification, but is based on sequence
XX information supplied by the European Patent Office
XX

Seq Sequence 7145 BP; 1896 A; 177 C; 1529 G; 3543 T; 0 U; 0 Other;
Query Match 87.2%; Score 21.8; DB 6; Length 7145;
Best Local Similarity 92.0%; Pred. No. 4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCGCAACAATCT 25
DB 1073 AAAAAAAAAATCGCAACAATCT 1049

RESULT 11
AAD22321/C
ID AAD22321 standard; DNA; 7145 BP.
XX
XX AAD22321;
XX
XX 12-FEB-2002 (first entry)
XX
XX
XX Chemically treated human genomic DNA #11 associated with DNA adducts.
XX
XX DNA adduct; peptide nucleic acid; PNA; cytosine methylation;
XX gene therapy; tumour; cancer; human; ds.
XX
XX Homo sapiens.
XX
XX WO200177378-A2.
XX
XX
XX 18-OCT-2001.
XX
XX
XX 06-APR-2001; 2001WO-EP004015.
XX

PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIC-) EPIDENOMICS AG.
 XX
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-010923/01.
 XX
 PT Novel nucleic acid comprising sequence of a segment of chemically
 PT pretreated DNA of genes associated with DNA adduct, useful for diagnosis
 PT and therapy of solid tumors and cancer.
 XX
 XX
 PS Claim 1; Page 52-53; 97pp; English.
 CC The invention relates to chemically modified DNA of genes associated with
 CC DNA adducts, oligonucleotides and/or peptide nucleic acid (PNA) oligomers
 CC for detecting cytosine methylations as well as method for ascertaining
 CC genetic and/or epigenetic parameters of genes associated with DNA
 CC adducts. Oligomers of the invention coupled to a solid phase is useful
 CC for manufacturing an arrangement of different oligomers (array) fixed to
 CC a carrier material for analysing diseases associated with the methylation
 CC state of the CpG dinucleotides of chemically pretreated DNA of genes
 CC associated with DNA adduct. They are also useful for ascertaining genetic
 CC and/or epigenetic parameters for the diagnosis and/or therapy of existing
 CC diseases or the predisposition to specific diseases by analysing cytosine
 CC methylations. Sequences of the invention are useful for diagnosis and
 CC therapy of solid tumors and cancers. They are also useful in gene
 CC therapy. The present sequence is chemically pretreated human genomic DNA
 CC associated DNA adducts
 XX
 SQ Sequence 7145 BP; 1896 A; 177 C; 1529 G; 3543 T; 0 U; 0 Other;
 Query Match 87.2%; Score 21.8; DB 6; Length 7145;
 Best Local Similarity 92.0%; Pred. No. 4e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAAATCT 25
 Db 1073 AAAAAAAAAATCGCAACAAATCT 1049
 RESULT 12
 ID ABL32095/c
 XX ABL32095 standard; DNA; 7928 BP.
 AC ABL32095;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 68.
 XX
 KM Human; immune system disease; cytosine methylation; antiaethmatic;
 KM antiatherosclerotic; antianaemic; cytosatic; noctropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antineumatic; antiaethritic; antidiabetic; antipsoriasis;
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KM ds.
 KW
 XX
 OS Homo sapiens.
 XX
 PN MO200200926-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.

XX
 PA (EPIC-) EPIDENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 XX
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 68; 32pp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 7928 BP; 1959 A; 198 C; 2043 G; 3728 T; 0 U; 0 Other;
 Query Match 87.2%; Score 21.8; DB 6; Length 7928;
 Best Local Similarity 92.0%; Pred. No. 4e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAAATCT 25
 Db 3542 AAAAAAAAAATCGCAACAAATCT 3518
 RESULT 13
 ID ABR31169/c
 XX ABR31169 standard; DNA; 7928 BP.
 AC ABR31169;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified complementary DNA #6.
 XX
 KM Human; signal transduction associated gene; cytosine methylation state;
 KM CpG island; signal transduction associated disease; solid tumour; cancer;
 KM antitumour; cytosatic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS
 PN WO200200926-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP007472.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIC-) EPIDENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-147896/19.
 XX
 XX
 PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction.
 XX
 PS Claim 1; SEQ ID NO 12; 24pp; English.
 CC The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically

CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or RNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 7928 BP; 1959 A; 198 C; 2043 G; 3728 T; 0 U; 0 Other;

Query Match 87.2%; Score 21.8; DB 6; Length 7928;
Best Local Similarity 92.0%; Pred. No. 4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATCT 25
DB 3542 AAAAAAAAAATCGCAACCAATCT 3518

RESULT 14
ABL70124/C
ID ABL70124 standard; DNA; 7928 BP.

AC ABL70124;

DT 01-JUN-2002 (first entry)

DE Chemically treated cell signalling DNA sequence complementary to #7.

KM Cell signalling; cytosine methylation; cell signalling disease; cancer;
KM tumour; cytosine; ds.

OS Unidentified.

PN WO200202807-A2.

PD 10-JAN-2002.

PF 29-JUN-2001; 2001WO-EP007471.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR MPI, 2002-154758/20.

PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling.

PS Claim 1; SEQ ID NO 14; 24pp + Sequence listing; English.

CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or RNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and

CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 7928 BP; 1959 A; 198 C; 2043 G; 3728 T; 0 U; 0 Other;

Query Match 87.2%; Score 21.8; DB 6; Length 7928;
Best Local Similarity 92.0%; Pred. No. 4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATCT 25
DB 3542 AAAAAAAAAATCGCAACCAATCT 3518

RESULT 15
AAS61056/C
ID AAS61056 standard; DNA; 7928 BP.

AC AAS61056;

DT 29-JAN-2002 (first entry)

DE Human gene regulation-associated gene oligonucleotide #11.

KM Human; Gene regulation-associated gene; severe combined immunodeficiency;
KM cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KM asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KM renal disease; Preeclampsia; cardiac allograft vascular disease;
KM colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KM immunostimulant; cardiac; anti-inflammatory; coagulant; antiasthmatic;
KM nephrotropic; gynecological; anti-tumour; immunosuppressive; cytosine.

OS Homo sapiens.

PN WO200177375-A2.

PD 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP003968.

PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR MPI, 2002-017470/02.

PT New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.

PS Claim 1; SEQ ID NO 12; 26pp; English.

CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are

PR	26-JUL-2000	2000US-0220964P
PR	14-AUG-2000	2000US-0224518P
PR	14-AUG-2000	2000US-0224519P
PR	14-AUG-2000	2000US-0225213P
PR	14-AUG-2000	2000US-0225213P
PR	14-AUG-2000	2000US-0225266P
PR	14-AUG-2000	2000US-0225267P
PR	14-AUG-2000	2000US-0225268P
PR	14-AUG-2000	2000US-0225270P
PR	14-AUG-2000	2000US-0225271P
PR	14-AUG-2000	2000US-0225757P
PR	14-AUG-2000	2000US-0225758P
PR	14-AUG-2000	2000US-0225759P
PR	12-AUG-2000	2000US-0226681P
PR	22-AUG-2000	2000US-0226868P
PR	22-AUG-2000	2000US-0227182P
PR	30-AUG-2000	2000US-0228709P
PR	01-SEP-2000	2000US-0229287P
PR	01-SEP-2000	2000US-0229343P
PR	01-SEP-2000	2000US-0229344P
PR	05-SEP-2000	2000US-0229345P
PR	05-SEP-2000	2000US-0229509P
PR	05-SEP-2000	2000US-0229513P
PR	06-SEP-2000	2000US-0230437P
PR	06-SEP-2000	2000US-0230438P
PR	08-SEP-2000	2000US-0231242P
PR	08-SEP-2000	2000US-0231243P
PR	08-SEP-2000	2000US-0231244P
PR	08-SEP-2000	2000US-0231413P
PR	08-SEP-2000	2000US-0231414P
PR	08-SEP-2000	2000US-0232080P
PR	08-SEP-2000	2000US-0232081P
PR	12-SEP-2000	2000US-0231968P
PR	14-SEP-2000	2000US-0232237P
PR	14-SEP-2000	2000US-0232338P
PR	14-SEP-2000	2000US-0232339P
PR	14-SEP-2000	2000US-0232400P
PR	14-SEP-2000	2000US-0232401P
PR	14-SEP-2000	2000US-0233063P
PR	14-SEP-2000	2000US-0233064P
PR	21-SEP-2000	2000US-0233065P
PR	21-SEP-2000	2000US-0234223P
PR	21-SEP-2000	2000US-0234274P
PR	25-SEP-2000	2000US-0234997P
PR	25-SEP-2000	2000US-0234998P
PR	25-SEP-2000	2000US-0235464P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235835P
PR	29-SEP-2000	2000US-0236337P
PR	29-SEP-2000	2000US-0236367P
PR	29-SEP-2000	2000US-0236368P
PR	29-SEP-2000	2000US-0236369P
PR	29-SEP-2000	2000US-0236370P
PR	29-SEP-2000	2000US-0236802P
PR	02-OCT-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	02-OCT-2000	2000US-0237040P
PR	13-OCT-2000	2000US-0239393P
PR	13-OCT-2000	2000US-0239397P
PR	20-OCT-2000	2000US-0240960P
PR	20-OCT-2000	2000US-0241121P
PR	20-OCT-2000	2000US-0241125P
PR	20-OCT-2000	2000US-0241785P
PR	20-OCT-2000	2000US-0241786P
PR	20-OCT-2000	2000US-0241787P
PR	20-OCT-2000	2000US-0241808P
PR	20-OCT-2000	2000US-0241809P
PR	01-NOV-2000	2000US-0241826P
PR	08-NOV-2000	2000US-0244617P
PR	08-NOV-2000	2000US-0246474P
PR	08-NOV-2000	2000US-0246475P

PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249279P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
PA	(HUMA-)	HUMAN GENOME SCI INC.
PI	Rosen CA, Barash SC, Ruben SM,	
DR	WPI; 2001-483426/52.	
XX		
PT	Nucleic acids encoding human immune/haematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and metastasis.	
XX		
PS	Disclosure; SEQ ID NO 40951; 3071pp + Sequence Listing; English.	
XX		
CC	AAK64951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
CC	amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I), by inserting the	
CC	nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/haematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/haematopoietic antigen genomic	


```

CC sequences from the present invention. AK54942 to AK54950 and AAW82246
CC represent sequences used in the exemplification of the present invention
XX
XX Sequence 37736 BP; 9869 A; 9250 C; 9209 G; 9408 T; 0 U; 0 Other;
QY
Query Match 87.2%; Score 21.8; DB 4; Length 37736;
Best Local Similarity 92.0%; Pred. No. 4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Dh 1 AAAAAAAAAATCGCAACAATCT 25
5567 AAAAAAAAAATCGCAACAATATCT 5591
RESULT 18
ABD33268/c
XX ID ABD33268 standard; DNA; 180227 BP.
XX
XX ABD33268;
XX AC
XX DT 18-NOV-2004 (first entry)
XX DE Human cancer-associated (CA) gene HD07-046.
XX KM Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX ds; cancer; cytostatic.
XX OS Homo sapiens.
XX PN WO2004058146-A2.
XX PD 15-JUL-2004.
XX PF 15-DEC-2003; 2003WO-US040081.
XX PR 17-DEC-2002; 2002US-00322281.
XX PA (SAGR-) SAGRES DISCOVERY INC.
XX PI Morris DW, Malandro MS;
XX WP; 2004-499109/47.
DR Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
PS Claim 16; SEQ ID NO 308; 182bp; English.
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_ptc_sequences
XX
SQ Sequence 180227 BP; 51503 A; 33645 C; 36852 G; 55474 T; 0 U; 2753 Other;
QY
Query Match 87.2%; Score 21.8; DB 13; Length 180227;
Best Local Similarity 92.0%; Pred. No. 3.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 AAAAAAAAAATCGCAACAATCT 25

```

DB	152244	AAAAAAAAAATCGCAAAAAAACT	152220
<hr/>			
RESULT 19			
ID	ABLJ34385/c		
XX	ABLJ34385 standard; DNA; 1710 BP.		
AC	ABLJ34385;		
DT	26-MAR-2002 (first entry)		
DE	Human immune system associated gene SEQ ID NO: 2358.		
XX			
KW	Human, immune system disease; cytosine methylation; antiasthmatic;		
KM	antiartherosclerotic; antianaemic; cytostatic; noctropic;		
KM	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;		
KM	anthrpaumatic; antiarthritic; antidiabetic; antisporadic;		
KM	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;		
KM	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;		
KM	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;		
ds.			
OS	Homo sapiens.		
PN	WO200200928-A2.		
PD	03-JAN-2002.		
PF	02-JUL-2001; 2001WO-EP007537.		
PR	30-JUN-2000; 2000DE-01032529.		
PR	01-SEP-2000; 2000DE-01043826.		
PA	(EPIG-) EPIGENOMICS AG.		
PI	Olek A, Piepenbrock C, Berlin K;		
DR	WPI; 2002-130909/17.		
PT	Nucleic acid comprising fragment of chemically modified gene, useful for		
PT	diagnosis and treatment of diseases associated with abnormal cytosine		
PT	methylation.		
PS	Claim 1; SEQ ID NO 2358; 32pp + Sequence Listing; German.		
CC	The present invention provides a number of human immune system associated		
CC	genes which are modified by the methylation of cytosines. The sequences		
CC	can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neovascular glaucoma and		
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid		
CC	leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,		
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel		
CC	diseases. The present sequence is a gene of the invention		
SEQ	Sequence 1710 BP; 378 A; 71 C; 433 G; 828 T; 0 U; 0 Other;		
<hr/>			
Query Match	85.6% Score 21.4; DB 6; Length 1710;		
Best Local Similarity	95.7% Pred. NO.5.Se+02;		
Match 22; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
Cy	1 AAAAAAAAAATCGCAAAACAAT 23		
Db	469 AAAAAAAAAAACGCCAACCAAT 447		
<hr/>			
RESULT 20			
ADSS31291			
ID	ADSS31291 standard; DNA; 230 BP.		
XX			
AC	ADSS31291;		
XX			
DT	18-NOV-2004 (first entry)		
XX			

DE Human genome high complexity repeat found in the HIRA gene #324.
 XX
 KM Human, ds;
 KM histone cell cycle regulation defective, S. cerevisiae homologue A, HIRA;
 KM high complexity repeat; in situ hybridization; Southern blot;
 KM chromosome breakpoint; inherited genetic disease; neoplastic disorder;
 KM chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.
 XX
 OS Homo sapiens.
 PN US200324356-A1.
 PD 04-DEC-2003.
 XX
 XX 14-MAY-2001; 2001US-00854867.
 PF 16-MAY-2000; 2000US-00573080.
 PR (KNOL/) KNOLL J H M.
 PA (ROGA/) ROGAN P K.
 XX
 PI Knoll JHM, Rogan PK;
 XX WPI; 2002-062378/08.
 DR
 XX
 PT Single copy genomic hybridization probes for detecting specific nucleic
 PT acid sequences in sample by in situ hybridization useful for detection of
 PT acquired or inherited genetic diseases.
 XX
 PS Example 1; SEQ ID NO 324; 30pp; English.
 XX
 XX The invention relates to a nucleic acid hybridization probe comprising a
 CC labelled, single copy nucleic acid of at least 50 nucleotides, which
 CC will hybridize to a deduced single copy sequence interval in target
 CC nucleic acid (TNA) of known sequence. The single copy sequence is deduced
 CC by comparing the target nucleic acid (e.g. a disease causing gene) with a
 CC collection of high and low complexity repeat sequences as found in the
 CC genome of the organism from containing the target nucleic acid. The probe
 CC is generated by PCR on the target sequence. The probe is essentially free
 CC of blocking nucleic acid sequences which will hybridize to repeat
 CC sequences within the genome of which the TNA is a part, and is labelled
 CC with a label selected from fluorochrome-responsive labels, fluorochromes,
 CC calorimetric chemical, conjugated proteins, antibodies, antigens and
 CC their mixtures. The probe is useful in a hybridization method, where the
 CC hybridization method is from in situ hybridization, Southern blot, and
 CC other methods in which nucleic acid is immobilized, where the method
 CC further comprises selecting a single copy nucleic acid which will
 CC hybridize to a duplilon or triplicon sequence domain. The probe is useful
 CC for determining the existence of previously unknown repeat sequence
 CC families in a genome. The method comprises reacting a labelled probe with
 CC the genome, causing the probe to hybridize and ascertaining if the probe
 CC hybridizes to the genome at more than three preferably ten different
 CC locations as a determination of new repeat sequence family, where the
 CC determining step comprises selecting the single copy sequence from a
 CC duplilon or triplicon sequence domain. The probe is useful for
 CC determining a chromosome breakpoint and is useful in the fields for
 CC cytogenetics and molecular genetics for determining the presence of
 CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
 CC the probes may be used to analyse specific chromosomal locations by in
 CC situ hybridization as a detection of acquired or inherited genetic
 CC diseases especially for detection of genetic or neoplastic disorders.
 CC Unlike prior art techniques, the probe permits more precise chromosomal
 CC breakpoint determinations by in situ hybridization. The genomic sequence
 CC comprising the human HIRA gene (histone cell cycle regulation defective,
 CC S. cerevisiae, homologue A) was analysed for single copy sequence
 CC intervals for use as probes of the invention. HIRA is located on
 CC chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and
 CC Velo-Cardio-facial syndromes. The present sequence is a high complexity
 CC repeat found within the human genome used to analyse the HIRA gene for
 CC repeat regions. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030224356.
 XX

SEQ Sequence 230 BP; 70 A; 51 C; 55 G; 51 T; 0 U; 3 Other;
 Query Match 84.0%; Score 21; DB 7; Length 230;
 Best Local Similarity 84.0%; Pred. No. 7.6e+02;
 Matches 21; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 AAAAAAAAAATCCGAACCAATCT 25
 Db 109 AAAAAAAAAATCCGAACCAATCT 133
 RESULT 21
 ADY36679
 ID ADY36679 standard; DNA; 230 BP.
 XX
 AC ADY36679;
 XX
 DT 05-MAY-2005 (first entry)
 XX
 DE HIRA genomic fragment SEQ ID NO 324.
 XX
 KM hybridization; DNA detection; neoplasm; genetic disorder; cytogenetics;
 KM HIRA; ds.
 XX
 OS Homo sapiens.
 PN WO200188089-A2.
 XX
 PD 22-NOV-2001.
 XX
 PF 15-MAY-2001; 2001WO-US015674.
 PR 16-MAY-2000; 2000US-00573080.
 PR 14-MAY-2001; 2001US-00854867.
 XX
 PA (CHIL-) CHILDREN'S MERCY HOSPITAL.
 PI Knoll JHM, Rogan PK, Casarzo PM;
 XX WPI; 2002-062378/08.
 DR
 XX
 PT Single copy genomic hybridization probes for detecting specific nucleic
 PT acid sequences in sample by in situ hybridization useful for detection of
 PT acquired or inherited genetic diseases.
 XX
 PS Example 1; SEQ ID NO 324; 67pp; English.
 XX
 XX The invention describes a nucleic acid hybridization probe (I) comprising
 CC a labelled, single copy nucleic acid of at least 50 nucleotides, which
 CC will hybridize to a deduced single copy sequence interval in target
 CC nucleic acid (TNA) of known sequence. (I) is useful in a hybridization
 CC method which comprises preparing a reaction mixture comprising TNA and
 CC (I) which hybridizes to TNA, and causing (I) to hybridize to TNA, where
 CC the hybridization method is from in situ hybridization, Southern blot,
 CC and other methods in which nucleic acid is immobilized, where the method
 CC further comprises selecting a single copy nucleic acid which will
 CC hybridize to a duplilon or triplicon sequence domain. (I) is useful for:
 CC determining the existence of previously unknown repeat sequence families
 CC in a genome; determining a chromosome breakpoint and in the fields of
 CC cytogenetics and molecular genetics for determining the presence of
 CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
 CC the probes may be used to analyse specific chromosomal locations by in
 CC situ hybridization as a detection of acquired or inherited genetic
 CC diseases especially for detection of genetic or neoplastic disorders.
 CC Unlike prior art techniques, (I) permits more precise chromosomal
 CC breakpoint determinations by in situ hybridization. Hybridization
 CC techniques utilizing (I), have made it possible to obtain reliable,
 CC easily detectable signals with relatively small probes. A readily
 CC detectable signal was obtained with a probe on the order of 2 kb in
 CC length, using fluorescent in situ hybridization (FISH) technology. This
 CC sensitivity of (I) is improved compared to the prior art, because the
 CC probes of (I) are homogeneous single copy sequences. However, smaller
 CC amplified segments, each comprising non-repetitive sequences, may also be

CC used in combination as probes to achieve adequate signals for in situ
 CC hybridization. Complex single copy probes that hybridize to duplicated or
 CC triplicated targets can also increase hybridization signals. This
 CC sequence represents a human HIRA genomic sequence that shows homology to
 CC a known high-complexity repeat sequence family of the human genome and is
 CC used in the creation of an HIRA gene probe.

XX Sequence 230 BP; 70 A; 51 C; 55 G; 51 T; 0 U; 3 Other;

Query Match 84.0%; Score 21; DB 7; Length 230;
 Best Local Similarity 84.0%; Pred. No. 7.6e+02;
 Matches 21; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATCT 25
 |||||:|||||:|||||
 DB 109 AAAAAAAAAATCGCAACCAATCT 133

RESULT 22

ID ABQ25827
 ABQ25827 standard; DNA; 544 BP.

AC ABQ25827;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 12418.

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-BP010074.

PR 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful for

PT diagnosis and prognosis, comprises selective hybridization of amplicons

PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridized to two classes, each with at least one member,

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

CC degree of hybridization to both classes is determined from the label on

CC the amplicon. From the ratio of labels hybridized to the two classes of

CC oligomers, the degree of methylation is calculated. The method is used:

CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs

CC and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,

CC particularly by detecting mutations or single nucleotide polymorphisms

CC (SNP's); and (ii) for differentiation of cell or tissue types and for

CC investigating cell differentiation. The method allows the methylation

CC status of many C residues to be determined simultaneously. ABQ13410-

CC ABQ54121 represent genomic DNA sequences used to illustrate the method

CC for determining the degree of cytosine methylation described in the

CC disclosure of the invention

XX Sequence 544 BP; 298 A; 143 C; 79 G; 20 T; 0 U; 4 Other;

Query Match 83.2%; Score 20.8; DB 6; Length 544;
 Best Local Similarity 91.7%; Pred. No. 8.7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATC 24
 |||||:|||||:|||||
 DB 133 AAAAAAAAAATCGCAACCAATC 156

RESULT 23

ID ABQ25826/c
 ABQ25826 standard; DNA; 544 BP.

AC ABQ25826;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 12417.

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-BP010074.

PR 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful for

PT diagnosis and prognosis, comprises selective hybridization of amplicons

PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridized to two classes, each with at least one member,

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

CC degree of hybridization to both classes is determined from the label on

CC the amplicon. From the ratio of labels hybridized to the two classes of

CC oligomers, the degree of methylation is calculated. The method is used:

CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs

CC and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,

CC particularly by detecting mutations or single nucleotide polymorphisms

CC (SNP's); and (ii) for differentiation of cell or tissue types and for

CC investigating cell differentiation. The method allows the methylation

CC status of many C residues to be determined simultaneously. ABQ13410-

CC disclosure of the invention

```
XX      SQ      Sequence 544 BP; 20 A; 79 C; 143 G; 298 T; 0 U; 4 Other;
Query Match      83.2%; Score 20.8; DB 6; Length 544;
Best Local Similarity 91.7%; Pred.No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1  AAAAAAAAAATCGCAACCAATC 24
         |||||
Db      412 AAAAAAAAAACGCAACCAACC 389

RESULT 24
AKS6935
ID      AAKS6935 standard; cDNA; 655 BP.
XX
AC      AAKS6935;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:1995.
XX
KW      Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX      cytostatic; gene therapy; vaccine; metastasis; ss.
XX      Homo sapiens.
XX      MO200157182-A2.
XX      09-AUG-2001.
PD
XX      17-JAN-2001; 2001WO-US001354.
PF
XX      31-JAN-2000; 2000US-0179065P.
PR      04-FEB-2000; 2000US-0180628P.
PR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205515P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      11-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225477P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226686P.
PR      22-AUG-2000; 2000US-0227182P.
PR      23-AUG-2000; 2000US-0227009P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      01-SEP-2000; 2000US-0229345P.
```

```
PR      05-SEP-2000; 2000US-0229509P.
PR      05-SEP-2000; 2000US-0229513P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231249P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0231415P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      14-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235834P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241121P.
PR      20-OCT-2000; 2000US-0241785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
PR      01-NOV-2000; 2000US-0244617P.
PR      08-NOV-2000; 2000US-0246474P.
PR      08-NOV-2000; 2000US-0246475P.
PR      08-NOV-2000; 2000US-0246476P.
PR      08-NOV-2000; 2000US-0246477P.
PR      08-NOV-2000; 2000US-0246478P.
PR      08-NOV-2000; 2000US-0246523P.
PR      08-NOV-2000; 2000US-0246524P.
PR      08-NOV-2000; 2000US-0246525P.
PR      08-NOV-2000; 2000US-0246526P.
PR      08-NOV-2000; 2000US-0246527P.
PR      08-NOV-2000; 2000US-0246528P.
PR      08-NOV-2000; 2000US-0246532P.
PR      08-NOV-2000; 2000US-0246533P.
PR      08-NOV-2000; 2000US-0246609P.
PR      08-NOV-2000; 2000US-0246610P.
PR      08-NOV-2000; 2000US-0246611P.
PR      08-NOV-2000; 2000US-0246613P.
PR      17-NOV-2000; 2000US-0249207P.
PR      17-NOV-2000; 2000US-0249208P.
PR      17-NOV-2000; 2000US-0249209P.
PR      17-NOV-2000; 2000US-0249210P.
PR      17-NOV-2000; 2000US-0249211P.
PR      17-NOV-2000; 2000US-0249212P.
PR      17-NOV-2000; 2000US-0249213P.
PR      17-NOV-2000; 2000US-0249214P.
PR      17-NOV-2000; 2000US-0249215P.
```

PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0251989P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251900P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-483426/52.
 XX P-PSDB; AAM84154.
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 PS
 XX
 PS Claim 1; SEQ ID NO 1995; 3071pp + Sequence listing; English.
 CC
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 655 BP; 236 A; 115 C; 100 G; 201 T; 0 U; 3 Other;
 QY
 Query Match 83.2%; Score 20.8; DB 4; Length 655;
 Best Local Similarity 91.7%; Pred. No. 8.7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 11 AAAAAAAAAAATGCAAAAAATC 24
 11 AAAAAAAAAAATGCAAAAAATC 34
 RESULT 25
 ID ABO35161 standard; DNA; 722 BP.
 XX ABO35161;
 AC
 XX
 DT 12-JUL-2002 (first entry)
 XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21752.
 XX
 XX Human; "Cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 PF
 PF 01-SEP-2001; 2001WO-EP010074.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPIC-) EPITENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 DR WPI: 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 XX from chemically treated DNA.
 XX
 PS Claim 12; 56pp + Sequence listing; 56pp; German.
 CC
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABO13410-
 CC ABO54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 722 BP; 410 A; 147 C; 54 G; 111 T; 0 U; 0 Other;
 QY
 Query Match 83.2%; Score 20.8; DB 6; Length 722;
 Best Local Similarity 91.7%; Pred. No. 8.7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 665 AAAAAAAAAAATGCAAAACTTAATC 688
 1 AAAAAAAAAAATGCAAAACTTAATC 24
 1 AAAAAAAAAAATGCAAAACTTAATC 688
 RESULT 26
 ID ABO35160/c
 XX ABO35160 standard; DNA; 722 BP.
 XX ABO35160;
 AC
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21751.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KM drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 OS Homo sapiens.
 XX
 XX WO200218632-A2.
 PN
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI, 2002-371829/40.
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 PS
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ5412 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 CC
 CC
 SQ Sequence 722 BP; 111 A; 54 C; 147 G; 410 T; 0 U; 0 Other;
 XX
 XX
 Query Match 83.2%; Score 20.8; DB 6; Length 722;
 Best Local Similarity 91.7%; Pred. No. 8.7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAATC 24
 Db 58 AAAAAAAAAATCGCAACAATC 35

XX
 OS Homo sapiens.
 XX
 XX WO200218632-A2.
 PN
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI, 2002-371829/40.
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 PS
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ5412 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 CC
 CC
 SQ Sequence 730 BP; 45 A; 57 C; 145 G; 483 T; 0 U; 0 Other;
 XX
 XX
 Query Match 83.2%; Score 20.8; DB 6; Length 730;
 Best Local Similarity 91.7%; Pred. No. 8.7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAATC 24
 Db 560 AAAAAAAAAATCGCAACAATC 537

PN WO200218632-A2.
XX 07-MAR-2002.
PD
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40;
DR
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert in a
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNPs); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 730 BP; 483 A; 145 C; 57 G; 45 T; 0 U; 0 Other;

Query Match 83.2%; Score 20.8; DB 6; Length 730;
Best Local Similarity 91.7%; Pred. No. 8.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATC 24
DB 171 AAAAAAAAAAAGCAAAAAAAAAATC 194

RESULT 29
AAK69138
ID AAK69138 standard; DNA; 1096 BP.
XX
AC AAK69138;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23950.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205155P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226281P.
PR 22-AUG-2000; 2000US-0226586P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231245P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232402P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234297P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.

CC encode promoter components of the cellular activation and transcription
CC apparatus (5' cis regulatory DNA sequences). The sequences may be used in
CC the modification of gene activation and/or expression in eukaryotes,
CC especially rice and other monocots. They have many applications in
CC standard molecular biology, e.g. as for chromosome and gene mapping, in
CC polymerase chain reaction (PCR) technology, for the study of gene
CC function and expression in vivo, specific down-regulation of target
CC genes, and in the production of sense and/or antisense nucleic acids. The
CC present sequence represents the coding sequence of rice GAPDH as
CC described in the invention
XX
SQ Sequence 1834 BP; 365 A; 473 C; 447 G; 478 T; 0 U; 71 Other;

Query Match 83.2%; Score 20.8; DB 6; Length 1834;
Best Local Similarity 91.7%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGCAACCAATC 24
DB 1626 AAAAAAAAAATCGCAACCAATC 1603

RESULT 31
ACL38225/c
ID ACL38225 standard; cDNA; 2000 BP.
XX
AC ACL38225;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:16788.
XX
XX ss; abiotic stress tolerance; transgenic plant; plant; cereal;
XX agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019668.
XX
PF 22-JUN-2001; 2001US-0300112P.
XX
PR 24-AUG-2001; 2001US-0314662P.
XX
PR 26-SEP-2001; 2001US-0325277P.
XX
PR 21-NOV-2001; 2001US-0332132P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PA Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F,
XX
PI Moughamer T, Provart N, Rieke D, Zhu T;
XX
XX MPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
XX
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
XX
PT stress, salt stress or osmotic stress.
XX
XX
PS Claim 48; SEQ ID NO 16788; 89pp; English.
XX
XX The invention relates to novel abiotic stress responsive polynucleotides
XX and polypeptides. Also disclosed are vectors, expression cassettes, host
XX cells, and plants containing such polynucleotides. Also disclosed are
XX methods for using the polynucleotides and polypeptides to alter the
XX responsiveness of a plant to abiotic stress. The invention is useful in
XX agriculture. The nucleic acid is useful for determining whether a test
XX plant has been exposed to an abiotic stress condition. It is also useful
XX for selecting an agent that alters abiotic stress regulated
XX polynucleotide expression in a plant cell, and to identify a homolog or
XX ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
XX molecule and the polypeptide encoded by it are useful in altering the
XX responsiveness of a plant to an abiotic stress, such as cold stress, salt

CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 489 A; 498 C; 431 G; 573 T; 0 U; 9 Other;

Query Match 83.2%; Score 20.8; DB 11; Length 2000;
Best Local Similarity 91.7%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGCAACCAATC 24
DB 1887 AAAAAAAAAATCGCAACCAATC 1864

RESULT 32
ACL38226
ID ACL38226 standard; cDNA; 2000 BP.
XX
AC ACL38226;
XX
DT 02-JUN-2005 (first entry)
XX
DE Rice stress-regulated promoter SEQ ID NO:16789.
XX
XX ss; abiotic stress tolerance; transgenic plant; plant; cereal;
XX agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
PD 30-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019668.
XX
PF 22-JUN-2001; 2001US-0300112P.
XX
PR 24-AUG-2001; 2001US-0314662P.
XX
PR 26-SEP-2001; 2001US-0325277P.
XX
PR 21-NOV-2001; 2001US-0332132P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PA Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F,
XX
PI Moughamer T, Provart N, Rieke D, Zhu T;
XX
XX MPI; 2003-248011/24.
XX
PT New stress-responsive nucleic acid, useful for altering the
XX
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
XX
PT stress, salt stress or osmotic stress.
XX
XX
PS Claim 48; SEQ ID NO 16789; 89pp; English.
XX
XX The invention relates to novel abiotic stress responsive polynucleotides
XX and polypeptides. Also disclosed are vectors, expression cassettes, host
XX cells, and plants containing such polynucleotides. Also disclosed are
XX methods for using the polynucleotides and polypeptides to alter the
XX responsiveness of a plant to abiotic stress. The invention is useful in
XX agriculture. The nucleic acid is useful for determining whether a test
XX plant has been exposed to an abiotic stress condition. It is also useful
XX for selecting an agent that alters abiotic stress regulated
XX polynucleotide expression in a plant cell, and to identify a homolog or
XX ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
XX molecule and the polypeptide encoded by it are useful in altering the
XX responsiveness of a plant to an abiotic stress, such as cold stress, salt
XX stress, osmotic stress or any of their combinations. The present sequence
XX is used in the exemplification of the invention
XX
SQ Sequence 2000 BP; 683 A; 472 C; 385 G; 452 T; 0 U; 8 Other;

Query Match 83.2%; Score 20.8; DB 11; Length 2000;
Best Local Similarity 91.7%; Pred. No. 8.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGAACAATC 24
 DB 1858 AAAAAAAAAATCGAACAATC 1881

RESULT 33
 ID ADJ40625 standard; cDNA; 2000 BP.
 AC ADJ40625;
 DT 06-MAY-2004 (first entry)
 DE Plant CDNA #1625.

XX Plant; gene; ss; transcription; plant genome augmentation; cereal;
 KM soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
 KM maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
 KM stress tolerance; salt tolerance; cold tolerance; drought tolerance;
 KM plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
 KM antifungal.

XX Eukaryota.
 OS
 XX US2004016025-A1.
 PN
 XX 22-JAN-2004.
 PD
 XX 26-SEP-2002; 2002US-00260238.
 PF
 XX 26-SEP-2001; 2001US-0325277P.
 PR 26-SEP-2001; 2001US-0325448P.
 PR 04-APR-2002; 2002US-0370620P.

XX (BUDW/) BUDWORTH P.
 PA (MOUG/) MOUGHAMER T.
 PA (BRIG/) BRIGGS S. P.
 PA (COOP/) COOPER B.
 PA (GLAZ/) GLAZEBROOK J.
 PA (GOLF/) GOLF S. A.
 PA (KATA/) KATAGIRI F.
 PA (KREP/) KREPS J.
 PA (PROV/) PROVART N.
 PA (RICK/) RIQUE D.
 PA (ZHUT/) ZHU T.

XX Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
 PI Goff SA, Katagiri F, Kreps J, Provart N, Rique D, Zhu T;
 XX WPI; 2004-190374/18.

XX New rice promoter, useful for manipulating crop plants to alter or
 PT improve phenotypic characteristics, e.g. produce large quantities of oil
 PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
 PT or high nutritional value.

XX Claim 26; SEQ ID NO 1625; 230bp; English.

XX The invention relates to plant nucleotide sequences that direct seed-,
 CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
 CC or constitutive transcription of an operatively linked nucleic acid
 CC segment. The invention also relates to a method for augmenting a plant
 CC genome and a method of identifying a gene, where its expression is
 CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
 CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
 CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
 CC sorghum, rice or wheat. The polynucleotides and the polypeptides they
 CC encode are useful for manipulating crop plants to alter or improve
 CC phenotypic characteristics, to produce large quantities of oil or
 CC proteins, to incur resistance to insecticides, viruses or fungi, and to
 CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
 CC have a high nutritional value with reduced apical dominance or dwarfism,

CC early flowering or altered metabolic pathways. This sequence represents a
 CC plant nucleic acid of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification but was obtained in
 CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 2000 BP; 489 A; 498 C; 431 G; 573 T; 0 U; 9 Other;

Query Match 83.2%; Score 20.8; DB 12; Length 2000;
 Best Local Similarity 91.7%; Pred. No. 8.6e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGAACAATC 24
 DB 1887 AAAAAAAAAATCGAACAATC 1864

RESULT 34
 ID ABL32182/c
 ABJ32182 standard; DNA; 5763 BP.

XX ABL32182;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Human immune system associated gene SEQ ID NO: 155.
 DE
 XX Human, immune system disease; cytosine methylation; antileukemic;
 KM anticancer; osteocytic; antianemic; cytosolic; noctropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; antiarthritic; antidiabetic; antiparasitic;
 KM antineoplastic; cancer; eye disease; arteriosclerosis; anemia;
 KM acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KM ds.

XX Homo sapiens.
 OS
 XX WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP007537.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PR (EPIC-) EPICENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX Claim 1; SEQ ID NO 155; 32bp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 5763 BP; 1629 A; 63 C; 1445 G; 2626 T; 0 U; 0 Other;

Query Match 83.2%; Score 20.8; DB 6; Length 5763;
 Best Local Similarity 91.7%; Pred. No. 8.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 AAAAAAAAAATCGCAACAATC 24
XX |||||
DE 1443 AAAAAAAAAAACCAACAATC 1420
XX

RESULT 35
ABL33615/C
ID ABL33615 standard; DNA; 6127 BP.
XX
AC ABL33615;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1588.
XX
KW Human; immune system disease; cytosine methylation; antiaethmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1588; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6127 BP; 1807 A; 129 C; 1286 G; 2905 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 20.8; DB 6; Length 6127;
Best Local Similarity 91.7%; Pred. No. 8.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACAATC 24
XX |||||
DB 494 AAAAAAAAAAACCAACAATC 471
XX
RESULT 36
AAS46503/C
ID AAS46503 standard; DNA; 6221 BP.
XX
AC AAS46503;
XX

```

```

DT 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #225.
DE
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
XX Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
XX
PR 06-APR-2000; 2000DE-01019056.
XX
PR 07-APR-2000; 2000DE-01019173.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 225; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumor suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumor suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6221 BP; 1610 A; 200 C; 1746 G; 2665 T; 0 U; 0 Other;
XX
Query Match 83.2%; Score 20.8; DB 4; Length 6221;
Best Local Similarity 91.7%; Pred. No. 8.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACAATC 24
XX |||||
DB 3559 AAAAAAAAAACGAACAATC 3536
XX
RESULT 37
ABL32795/C
ID ABL32795 standard; DNA; 8951 BP.
XX
AC ABL32795;
XX

```

DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 768.
 XX
 XX Human; immune system diseases; cytosine methylation; antiasthmatic;
 KM antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; antirachitic; antidiabetic; antipsoriatic;
 KM antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KM ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 768; 32bp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 SQ Sequence 8951 BP; 2118 A; 548 C; 2330 G; 3955 T; 0 U; 0 Other;
 XX
 Query Match 83.2%; Score 20.8; DB 6; Length 8951;
 Best Local Similarity 91.7%; Pred. No. 8.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCCGAACCAATC 24
 Db 3089 AAAAAAAAAACGCAAAAAAATC 3066
 XX
 RESULT 38
 ADS89440/C
 ID ADS89440 standard; DNA; 9859 BP.
 XX
 AC ADS89440;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 OS Oligonucleotide of the invention SEQ ID NO:456.
 XX
 KM ss; cell proliferative disorder; breast; methylation; cytosstatic;
 KM gene therapy; single nucleotide polymorphism; SNP.
 XX
 OS Unidentified.
 XX
 PN WO2004035803-A2.
 XX

PD 29-APR-2004.
 XX
 PF 01-OCT-2003; 2003WO-EP010881.
 XX
 PR 01-OCT-2002; 2002DE-01045779.
 XX
 PR 07-JAN-2003; 2003DE-01000096.
 XX
 PR 17-APR-2003; 2003DE-01017955.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Fockens J, Harbeck N, Koenig T, Maier S, Martens J, Model F;
 PI Nimrich I, Rujan T, Schmitt A, Schmitt M, Look MP, Marx A;
 XX
 DR WPI; 2004-348468/32.
 XX
 PT Predicting responsiveness of a subject with breast cell proliferative
 PT disorder, useful for treating or differentiating breast cell
 PT proliferative disorders comprises analyzing methylation pattern of a
 PT genomic DNA from the subject.
 XX
 PS Claim 25; SEQ ID NO 456; 104bp; English.
 XX
 CC The invention relates to a novel method for predicting the responsiveness
 CC of a subject with a cell proliferative disorder of the breast tissues to
 CC a therapy comprising analysing the methylation pattern of a target
 CC nucleic acid by contacting at least one of the target nucleic acids in a
 CC biological sample obtained from the subject prior to or during treatment.
 CC The method of the invention has cytosstatic activity, and may have a use
 CC in gene therapy. The set of oligonucleotides comprising at least two of
 CC the oligomers are useful for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms (SNPs) within the sequences. The
 CC method, nucleic acid, oligonucleotide, and kit are useful for the
 CC treatment, characterisation, classification and/or differentiation, of
 CC breast cell proliferative disorders. The method is also useful for
 CC predicting the responsiveness of a subject with a cell proliferative
 CC disorder of the breast tissues to a therapy. The present sequence is used
 CC in the exemplification of the invention.
 XX
 SQ Sequence 9859 BP; 2917 A; 171 C; 1923 G; 4848 T; 0 U; 0 Other;
 XX
 Query Match 83.2%; Score 20.8; DB 13; Length 9859;
 Best Local Similarity 91.7%; Pred. No. 8.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCCGAACCAATC 24
 Db 315 AAAAAAAAAATACCAACCAATC 292
 XX
 RESULT 39
 ADS89714/C
 ID ADS89714 standard; DNA; 9859 BP.
 XX
 AC ADS89714;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 OS Oligonucleotide of the invention SEQ ID NO:730.
 XX
 KM ss; cell proliferative disorder; breast; methylation; cytosstatic;
 KM gene therapy; single nucleotide polymorphism; SNP.
 XX
 OS Unidentified.
 XX
 PN WO2004035803-A2.
 XX
 PD 29-APR-2004.
 XX
 PF 01-OCT-2003; 2003WO-EP010881.
 XX
 PR 01-OCT-2002; 2002DE-01045779.
 XX
 PR 07-JAN-2003; 2003DE-01000096.
 XX
 PR 17-APR-2003; 2003DE-01017955.

XX (EPiG-) EPIGENOMICS AG.
PA Foekens J, Harbeck N, Koenig T, Maier S, Martens J, Model F;
XX Nimnich I, Rujan T, Schmitt A, Schmitt M, Look MP, Marx A;
PI WPI; 2004-348468/32.
XX
XX
XX Predicting responsiveness of a subject with breast cell proliferative
PT disorder, useful for treating or differentiating breast cell
PT proliferative disorders comprises analyzing methylation pattern of a
PT genomic DNA from the subject.
XX
XX Claim 25; SEQ ID NO 730; 104bp; English.
XX
XX The invention relates to a novel method for predicting the responsiveness
CC of a subject with a cell proliferative disorder of the breast tissues to
CC a therapy comprising analysing the methylation pattern of a target
CC nucleic acid by contacting at least one of the target nucleic acids in a
CC biological sample obtained from the subject prior to or during treatment.
CC The method of the invention has cytosstatic activity, and may have a use
CC in gene therapy. The set of oligonucleotides comprising at least two of
CC the oligomers are useful for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC methods, nucleic acid, oligonucleotide, and kit are useful for the
CC treatment, characterisation, classification and/or differentiation, of
CC breast cell proliferative disorders. The method is also useful for
CC predicting the responsiveness of a subject with a cell proliferative
CC disorder of the breast tissues to a therapy. The present sequence is used
CC in the exemplification of the invention.
XX
XX
SQ Sequence 9859 BP; 2917 A; 0 C; 1923 G; 5019 T; 0 U; 0 Other;
Query Match 83.2%; Score 20.8; DB 13; Length 9859;
Best Local Similarity 91.7%; Pred. No. 8.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACCAATC 24
DB 315 AAAAAAAAAATACCAACCAATC 292
RESULT 40
ABL49393/C
ID ABL49393 standard; DNA; 10144 BP.
XX
XX ABL49393;
AC
XX
XX 01-MAY-2002 (first entry)
DT
XX
XX Human polynucleotide associated with DNA replication SEQ ID NO 93.
DE
XX
XX Human; cytosstatic; neuroprotective; nootropic; immunostimulant;
KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
KW ERCC3; SNRPB; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;
KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO20017377-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 06-APR-2001; 2001WO-EP003971.
PF
XX
XX 06-APR-2000; 2000DE-01019058.
PR
XX 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI

XX WPI; 2002-017471/02.
DR
XX
XX New nucleic acid sequences from chemically modified genes associated with
PT DNA replication, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. Ataxia telangiectasia.
PT
XX
XX Claim 1; SEQ ID NO 93; 23bp; English.
PS
XX
XX The invention relates to nucleic acid sequences comprising at least 18
CC bases of a chemically pretreated gene associated with gene regulation,
CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPB,
CC RAD50 and LIG2. The chemical pretreatment converts cytosine bases
CC to methylated at the 5-position to uracil or another base with
CC hybridisation behaviour dissimilar to cytosine, to enable analysis of
CC cytosine methylations. The DNA sequences and method are useful in the
CC diagnosis of diseases (or predisposition to diseases) associated with DNA
CC replication and in therapy of such diseases, by enabling analysis of the
CC cytosine methylation patterns of such genes. They are especially useful
CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's
CC syndrome, solid tumours and cancer. Note: The sequence data for this
CC patent did not form part of the printed specification and was supposed to
CC be available directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences. However, the sequence data did
CC not correspond to that referred to in the specification. The present data
CC is taken from EPO data for the patent
XX
XX
SQ Sequence 10144 BP; 1983 A; 360 C; 2968 G; 4827 T; 0 U; 6 Other;
Query Match 83.2%; Score 20.8; DB 6; Length 10144;
Best Local Similarity 91.7%; Pred. No. 8.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACCAATC 24
DB 9717 AAAAAAAAAATCGTAACCAAC 9694
RESULT 41
ABL33275/C
ID ABL33275 standard; DNA; 13038 BP.
XX
XX ABL33275;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Human immune system associated gene SEQ ID NO: 1248.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosstatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antirhectic; antidiabetic; antiporotic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
XX Homo sapiens.
OS
XX
XX WO200200928-A2.
PN
XX
XX 03-JAN-2002.
PD
XX
XX 02-JUL-2001; 2001WO-EP007537.
PF
XX
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI

DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 1248; 32bp + Sequence Listing; German.
 CC
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 13038 BP; 3975 A; 86 C; 2547 G; 6430 T; 0 U; 0 Other;
 Query Match 83.2%; Score 20.8; DB 6; Length 13038;
 Best Local Similarity 91.7%; Pred. NO. 8.4e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACCAATC 24
 DB 5322 AAAAAAAAAATCGCAACCAATC 5299
 RESULT 42
 ABK31234/C
 ID ABK31234 standard; DNA; 14542 BP.
 AC
 XX ABK31234;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified DNA #39.
 XX
 KW Human; signal transduction associated gene; cytosine methylation state;
 KW CpG island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS
 PN WO200200926-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP007472.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-147896/19.
 DR
 XX
 PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction.
 XX
 PS Claim 1; SEQ ID NO 77; 24bp; English.
 CC
 CC The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
 CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
 CC the cytosine methylation state (CpG islands) of these genes, and a method
 CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
 CC genes associated with signal transduction. The genomic DNA can be

CC obtained from cells or cellular components which contain DNA, e.g. cell
 CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
 CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
 CC brain, heart, prostate, lung, breast or liver, histologic object slides,
 CC and all their possible combinations. The sequences of the invention are
 CC useful for the diagnosis and therapy of diseases associated with signal
 CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
 CC chemically pretreated genomic DNA sequences of different genes associated
 CC with signal transduction, or their complementary sequences. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office
 XX
 SQ Sequence 14542 BP; 2958 A; 332 C; 4239 G; 7013 T; 0 U; 0 Other;
 Query Match 83.2%; Score 20.8; DB 6; Length 14542;
 Best Local Similarity 91.7%; Pred. NO. 8.4e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACCAATC 24
 DB 7698 AAAAAAAAAACACAAACCAATC 7675
 RESULT 43
 ABL70191/C
 ID ABL70191 standard; DNA; 14542 BP.
 AC
 XX ABL70191;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Chemically treated cell signalling DNA sequence#41.
 XX
 KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
 KW tumour; cytostatic; ds.
 XX
 OS Unidentified.
 OS
 PN WO200202807-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP007471.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-154758/20.
 DR
 XX
 PT Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signaling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signaling.
 XX
 PS Claim 1; SEQ ID NO 81; 24bp + Sequence Listing; English.
 CC
 CC The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signaling. The activity of the modified sequence of the
 CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signaling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signaling. Note: The

CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 14542 BP; 2958 A; 332 C; 4239 G; 7013 T; 0 U; 0 Other;
Query Match 83.2%; Score 20.8; DB 6; Length 14542;
Best Local Similarity 91.7%; Pred. No. 8.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCGCAACCAATC 24
Db 7698 AAAAAAAAAAACACCAACCAATC 7675

RESULT 44
AAS61147/C
ID AAS61147 standard; DNA; 14542 BP.
XX
AC AAS61147;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #102.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotoxic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003968.
XX
PR 06-APR-2000; 2000DE-01019058.
XX
PR 07-APR-2000; 2000DE-01019173.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.
XX
PS Claim 1; SEQ ID NO 105; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/array) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes. Kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this

CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 14542 BP; 2958 A; 332 C; 4239 G; 7013 T; 0 U; 0 Other;
Query Match 83.2%; Score 20.8; DB 6; Length 14542;
Best Local Similarity 91.7%; Pred. No. 8.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCGCAACCAATC 24
Db 7698 AAAAAAAAAAACACCAACCAATC 7675

RESULT 45
AAS46519/C
ID AAS46519 standard; DNA; 15872 BP.
XX
AC AAS46519;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #241.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
XX
PR 06-APR-2000; 2000DE-01019058.
XX
PR 07-APR-2000; 2000DE-01019173.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 241; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC dinucleotide, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analyzing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences

PT Detecting and distinguishing colorectal cell proliferative disorders by
PT contacting genomic DNA of biological sample with reagent that
PT distinguishes methylated and non-methylated CpG dinucleotides within
PT target sequence of genomic DNA.
XX
XX
PS Claim 11; SEQ ID NO 293; 23pp; English.
XX
XX The invention relates to a novel method for detecting and distinguishing
CC between, or among, colorectal cell proliferative disorders. The method
CC involves contacting genomic DNA of a biological sample obtained from the
CC subject with one or more reagent(s), or a series of reagents that
CC distinguishes between methylated and non-methylated CpG dinucleotides
CC within a target sequence of the genomic DNA. The invention further
CC comprises: a nucleic acid comprising a sequence of 18 or more contiguous
CC nucleotides of a treated genomic DNA sequence chosen from any one of 284
CC fully defined nucleotide sequences, whose sequence listing is not
CC provided in the specification, and their complementary sequences, where
CC the contiguous sequence has one or more CpG, Tpa, or Cpa dinucleotide,
CC and the treatment is suitable to convert one or more of the unmethylated
CC cytosine base(s) of the genomic DNA sequence initially to uracil or
CC another base that is detectably dissimilar to cytosine in terms of
CC hybridization; an oligomer or peptide nucleic acid (PNA)-oligomer,
CC comprising 9 or more contiguous nucleotides that is complementary to or
CC hybridizes under moderately stringent or stringent conditions to one of
CC the 284 DNA sequences and their complementary sequences provided in the
CC source document, which is treated; a set of oligomers comprising two or
CC more of the oligomer of PNA-oligomer; an array of oligomers; and a kit
CC for carrying out the above methods. The method and its novel compositions
CC have cytosine activity. The polynucleotide sequence may be used in gene
CC therapy. The above methods are useful for detecting and distinguishing
CC between, or among, colorectal cell proliferative disorders chosen from
CC colorectal carcinoma, colon adenomas and colon polyps, in a biological
CC sample, such as histological slides, biopsies, paraffin embedded tissue,
CC bodily fluids, stool, blood, serum, plasma and their combinations. The
CC oligomer array is useful as a probe for detecting one or more of the
CC cytosine methylation state, or single nucleotide polymorphisms within the
CC genomic DNA or their complementary sequences. The polynucleotides of the
CC invention are useful for classifying, distinguishing between, or among,
CC diagnosing or determining the predisposition for colorectal cell
CC proliferative disorders, or for the therapy of colorectal cell
CC proliferative disorders. This polynucleotide sequence represents a DNA
CC fragment associated with the development of colon cell proliferative
CC disorders, used in the novel methylated/non-methylated CpG dinucleotide
CC detection method of the invention. Note: This sequence is not shown in
CC the specification. It has been electronically downloaded from the USPTO
CC website.
XX
XX
SQ Sequence 17968 BP; 3854 A; 0 C; 5867 G; 8247 T; 0 U; 0 Other;
XX
XX
Query Match 83.2%; Score 20.8; DB 14; Length 17968;
Best Local Similarity 91.7%; Pred. No. 8.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAACCAATC 24
Db 17916 AAAAAAAAAACCTCAACCAATC 17993
XX
XX
RESULT 48
AAL35943
ID AAL35943 standard; DNA; 32132 BP.
XX
XX AAL35943;
AC
XX
XX
DT 08-JAN-2002 (first entry)
XX
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 2308.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antineoplastic; anticancer;
XX vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX

KW musculoskeletal system; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200155367-A1.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001338.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214868P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0231415P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0233397P.
PR 14-SEP-2000; 2000US-0233398P.
PR 14-SEP-2000; 2000US-0233399P.
PR 14-SEP-2000; 2000US-0233400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 23-SEP-2000; 2000US-0234997P.
PR

PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
FA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-451937/48.
XX
PT Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the musculoskeletal system including musculoskeletal
PT cancers and also for testing and detection e.g. diagnosis.
XX
PS Example 2; SEQ ID NO 2308; 781bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAL07060-AAL07066) and proteins
CC (AAB03087-AB04109) associated with the musculoskeletal system useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. The genes are isolated from a range of human tissues
CC disclosed in the specification. The nucleic acids, proteins, antibodies
CC and (ant)agonists are useful in the diagnosis, treatment and prevention
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;
SQ
Query Match 83.2%; Score 20.8; DB 4; Length 32132;
Best Local Similarity 91.7%; Pred. No. 8.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCGCAACCAATC 24
Db 31617 AAAAAAAAAAGCAACCAATC 31640
RESULT 49
AAL07060
ID AAL07060 standard; DNA; 32132 BP.
XX
AC AAL07060;
XX
DT 21-NOV-2001 (first entry)
DE Human reproductive system related antigen DNA SEQ ID NO: 9748.
XX
XX Human reproductive system related antigen; reproductive system disorder;
KW Human; gene therapy; ds.
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.

PR	30-JUN-2000	2000US-02148866
PR	30-JUN-2000	2000US-02161357
PR	07-JUL-2000	2000US-02166477
PR	07-JUL-2000	2000US-02166807
PR	11-JUL-2000	2000US-02174887
PR	11-JUL-2000	2000US-02174867
PR	14-JUL-2000	2000US-02174966
PR	26-JUL-2000	2000US-02182620
PR	26-JUL-2000	2000US-02209630
PR	26-JUL-2000	2000US-02209649
PR	14-AUG-2000	2000US-02245188
PR	14-AUG-2000	2000US-02245189
PR	14-AUG-2000	2000US-02252133
PR	14-AUG-2000	2000US-02252144
PR	14-AUG-2000	2000US-02252666
PR	14-AUG-2000	2000US-02252677
PR	14-AUG-2000	2000US-02252689
PR	14-AUG-2000	2000US-02252799
PR	14-AUG-2000	2000US-02254477
PR	14-AUG-2000	2000US-02257588
PR	14-AUG-2000	2000US-02257589
PR	14-AUG-2000	2000US-02257599
PR	18-AUG-2000	2000US-02262799
PR	22-AUG-2000	2000US-02265619
PR	22-AUG-2000	2000US-02266868
PR	22-AUG-2000	2000US-02271828
PR	22-AUG-2000	2000US-02277099
PR	30-AUG-2000	2000US-02270099
PR	01-SEP-2000	2000US-02292877
PR	01-SEP-2000	2000US-02293437
PR	01-SEP-2000	2000US-02293438
PR	01-SEP-2000	2000US-02293439
PR	01-SEP-2000	2000US-02293440
PR	01-SEP-2000	2000US-02293441
PR	01-SEP-2000	2000US-02293442
PR	01-SEP-2000	2000US-02293443
PR	01-SEP-2000	2000US-02293444
PR	01-SEP-2000	2000US-02293445
PR	01-SEP-2000	2000US-02293446
PR	01-SEP-2000	2000US-02293447
PR	01-SEP-2000	2000US-02293448
PR	01-SEP-2000	2000US-02293449
PR	01-SEP-2000	2000US-02293450
PR	01-SEP-2000	2000US-02293451
PR	01-SEP-2000	2000US-02293452
PR	01-SEP-2000	2000US-02293453
PR	01-SEP-2000	2000US-02293454
PR	01-SEP-2000	2000US-02293455
PR	01-SEP-2000	2000US-02293456
PR	01-SEP-2000	2000US-02293457
PR	01-SEP-2000	2000US-02293458
PR	01-SEP-2000	2000US-02293459
PR	01-SEP-2000	2000US-02293460
PR	01-SEP-2000	2000US-02293461
PR	01-SEP-2000	2000US-02293462
PR	01-SEP-2000	2000US-02293463
PR	01-SEP-2000	2000US-02293464
PR	01-SEP-2000	2000US-02293465
PR	01-SEP-2000	2000US-02293466
PR	01-SEP-2000	2000US-02293467
PR	01-SEP-2000	2000US-02293468
PR	01-SEP-2000	2000US-02293469
PR	01-SEP-2000	2000US-02293470
PR	01-SEP-2000	2000US-02293471
PR	01-SEP-2000	2000US-02293472
PR	01-SEP-2000	2000US-02293473
PR	01-SEP-2000	2000US-02293474
PR	01-SEP-2000	2000US-02293475
PR	01-SEP-2000	2000US-02293476
PR	01-SEP-2000	2000US-02293477
PR	01-SEP-2000	2000US-02293478
PR	01-SEP-2000	2000US-02293479
PR	01-SEP-2000	2000US-02293480
PR	01-SEP-2000	2000US-02293481
PR	01-SEP-2000	2000US-02293482
PR	01-SEP-2000	2000US-02293483
PR	01-SEP-2000	2000US-02293484
PR	01-SEP-2000	2000US-02293485
PR	01-SEP-2000	2000US-02293486
PR	01-SEP-2000	2000US-02293487
PR	01-SEP-2000	2000US-02293488
PR	01-SEP-2000	2000US-02293489
PR	01-SEP-2000	2000US-02293490
PR	01-SEP-2000	2000US-02293491
PR	01-SEP-2000	2000US-02293492
PR	01-SEP-2000	2000US-02293493
PR	01-SEP-2000	2000US-02293494
PR	01-SEP-2000	2000US-02293495
PR	01-SEP-2000	2000US-02293496
PR	01-SEP-2000	2000US-02293497
PR	01-SEP-2000	2000US-02293498
PR	01-SEP-2000	2000US-02293499
PR	01-SEP-2000	2000US-02293500
PR	01-SEP-2000	2000US-02293501
PR	01-SEP-2000	2000US-02293502
PR		

PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244517P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249246P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249279P.
PR	17-NOV-2000;	2000US-0249289P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250316P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251033P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC,	Ruben SM;
XX		
DR	WPI; 2001-465570/50.	
XX		
PT	Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.	
XX		
PS	Disclosure; SEQ ID NO 9748; 1297bp + Sequence listing; English.	
XX		
CC	The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention	
CC		
CC		
CC		
XX	Sequence 32112 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;	

Query Match 83.2%; Score 20.8; DB 8; Length 32132;
Best Local Similarity 91.7%; Pred. No. 8.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACCAATC 24
Db 31617 AAAAAAAAAAGCAACCAATC 31640

Search completed: December 14, 2005, 02:42:18
Job time : 211.2 secs

THIS PAGE BLANK (USPTO)


```
96 20.8 83.2 604 2 B1178992 EST519937 B1178992 EST519937
97 20.8 83.2 605 2 CN006484 CSECS134E CN006484 CSECS134E
98 20.8 83.2 609 2 BG591621 EST499463 BG591621 EST499463
99 20.8 83.2 614 2 BE472121 EST416974 BE472121 EST416974
100 20.8 83.2 616 8 DN272002 L1B30321 DN272002 L1B30321
101 20.8 83.2 620 2 CN461909 10076.1 A CN461909 10076.1 A
102 20.8 83.2 624 2 BF460258 073D11 Ma BF460258 073D11 Ma
103 20.8 83.2 624 2 A0272138 nxbx0027F A0272138 nxbx0027F
104 20.8 83.2 625 1 AV705106 AV705106 AV705106
105 20.8 83.2 633 8 DN743595 94275.1 F DN743595 94275.1 F
106 20.8 83.2 636 7 CR636323 34888 In CR636323 34888 In
107 20.8 83.2 637 9 AZ654285 1M0528D24 AZ654285 1M0528D24
108 20.8 83.2 640 6 CB467475 733226 MA CB467475 733226 MA
109 20.8 83.2 640 8 DN634818 Cm mx0_39 DN634818 Cm mx0_39
110 20.8 83.2 641 8 DN551245 Cm mx0_32 DN551245 Cm mx0_32
111 20.8 83.2 662 7 CR850866 10876 Sco CR850866 10876 Sco
112 20.8 83.2 664 7 CR853996 14974 Sco CR853996 14974 Sco
113 20.8 83.2 666 2 B1175995 EST517090 B1175995 EST517090
114 20.8 83.2 675 6 CB418298 591090 MA CB418298 591090 MA
115 20.8 83.2 679 7 CR853280 14038 Sto CR853280 14038 Sto
116 20.8 83.2 687 7 CO892133 BOYGEN_20 CO892133 BOYGEN_20
117 20.8 83.2 690 6 CF179977 815023 MA CF179977 815023 MA
118 20.8 83.2 700 1 AW812825 RC3-ST018 AW812825 RC3-ST018
119 20.8 83.2 708 6 CF179662 814665 MA CF179662 814665 MA
120 20.8 83.2 709 9 B94518 CIT-HSP-216 B94518 CIT-HSP-216
121 20.8 83.2 731 9 AF010795 AF010795 AF010795
122 20.8 83.2 731 10 AG172059 AG172059 Pan trogl
123 20.8 83.2 746 7 CN216071 29927 Sub CN216071 29927 Sub
124 20.8 83.2 760 7 CO076049 GR_Ea37D CO076049 GR_Ea37D
125 20.8 83.2 777 7 CV435121 58644.1 S CV435121 58644.1 S
126 20.8 83.2 780 2 BG199198 RST18477 BG199198 RST18477
127 20.8 83.2 784 2 BG685290 602637724 BG685290 602637724
128 20.8 83.2 791 6 CB310977 AGENCOURT CB310977 AGENCOURT
129 20.8 83.2 798 7 CO886544 BOYGEN_14 CO886544 BOYGEN_14
130 20.8 83.2 807 8 DN045696 JGI_CABA4 DN045696 JGI_CABA4
131 20.8 83.2 811 8 CX968654 JGI_CAAP3 CX968654 JGI_CAAP3
132 20.8 83.2 820 7 CN212812 26333 Sub CN212812 26333 Sub
133 20.8 83.2 868 10 AG852749 Oryza sat AG852749 Oryza sat
134 20.8 83.2 879 9 A0364370 nxbx0060L A0364370 nxbx0060L
135 20.8 83.2 889 10 C1541962 OB_BA006 C1541962 OB_BA006
136 20.8 83.2 903 7 CV286646 66946.1 A CV286646 66946.1 A
137 20.8 83.2 926 9 CC623579 OGLBQ48TV CC623579 OGLBQ48TV
138 20.8 83.2 947 7 CK270765 EST116843 CK270765 EST116843
139 20.8 83.2 947 9 A0747328 HS_5536 A A0747328 HS_5536 A
140 20.8 83.2 958 5 BU146732 AGENCOURT BU146732 AGENCOURT
141 20.8 83.2 1006 9 CC301630 CH261-69K CC301630 CH261-69K
142 20.8 83.2 1129 10 C1104017 ISB1-42F1 C1104017 ISB1-42F1
143 20.8 83.2 1190 9 CC290361 CH261-60E CC290361 CH261-60E
144 20.8 83.2 1236 9 CC261436 C6751619 P046-1-DO CC261436 C6751619 P046-1-DO
145 20.8 83.2 1261 10 CG751619 AT203865 Tetradon CG751619 AT203865 Tetradon
146 20.8 83.2 175 10 CA402166 EL01N0432 CA402166 EL01N0432
147 20.4 81.6 236 6 A1374888 ILA0909.X A1374888 ILA0909.X
148 20.4 81.6 301 1 CK229948 TILUMIGEN CK229948 TILUMIGEN
149 20.4 81.6 314 7 BZ603936 WHAAG48TR BZ603936 WHAAG48TR
150 20.4 81.6 349 9
```

ALIGNMENTS

```
RESULT 1 424 bp DNA linear GSS 16-JUN-1998
LOCUS A0023140/c
DEFINITION HS_2177_A2_H07_MF CIT Approved Human Genomic Sperm Library D Homo
sequence.
Accession A0023140 GI:3220387
VERSION A0023140
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```

REFERENCE 1 (bases 1 to 424)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
COMMENT PUBMED 10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2177 row: O column: 14
Class: BAC ends
High quality sequence stop: 424.

FEATURES
source
Location/Qualifiers
1..424
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2177 Col=14 Row=O"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 93.6%, Score 23.4; DB 9; Length 424;
Best local Similarity 96.0%; Pred. No. 1.5e+03;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Gy 1 AAAAAAAAAATCCAAACAATCT 25
DB 101 AAAAAAAAAATCCAAACAATCT 77

```
RESULT 2 459 bp mRNA linear EST 03-JUN-2005
LOCUS B0929968/c
DEFINITION B0929968 Daphnia magna cDNA library Daphnia magna cDNA clone
IG0001_0051_A09.r.3, mRNA sequence.
Accession B0929968
VERSION B0929968
KEYWORDS Daphnia magna
SOURCE Daphnia magna
ORGANISM Daphnia magna; Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda;
Diplostecata; Cladocera; Anomopoda; Daphniidae; Daphnia.
REFERENCE 1 (bases 1 to 459)
AUTHORS Watanabe,H., Tatarazako,N., Oda,S., Nishide,H., Uchiyama,I., Morita,M. and Iguchi,T.
TITLE Analysis of Expressed Sequence Tags of the water flea, Daphnia magna
JOURNAL Genome (2005) In press
COMMENT Contact: Hajime Watanabe
Okazaki Institute for Bioscience
National Institute of Natural Sciences
Higashi-yama 5-1, Myodai-ji, Okazaki, Aichi, 444-8787, Japan
Tel: 81-564-59-5237
Fax: 81-564-59-5236
Email: watanabe@nib.ac.jp.
Location/Qualifiers
1..459
/organism="Daphnia magna"
/mol_type="mRNA"
/db_xref="taxon:35525"
/clone="IG0001_0051_A09.r"
/sex="female"
```



```

ORIGIN
Query Match          93.6% Score 23.4; DB 3; Length 459;
Best Local Similarity 96.0%; Pred No.1.5e+03;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGCAACAAATCT 25
    |||||
Db 277 AAAAAAAAAATCGCAACAAATCT 253
    |||||

RESULT 3
LOCUS B35642/c
DEFINITION HS-1029-A2-G01-MR.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT808 Col=2 Row=M, genomic survey sequence.
ACCESSION B35642
VERSION B35642
KEYWORDS B35642.1 GI:2535011
SOURCE GSS.
ORGANISM Homo sapiens (human)
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 492) Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Tractocoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
JOURNAL Unpublished (1997)
COMMENT Contact: Mahairas GG, Zackrone KD, Hood L University of Washington Seattle, WA 98195, USA Tel.: (206) 616-8744 Fax: (206) 685-7301 Email: kzackron@u.washington.edu
SEQUENCE Sequence Tagged Connector
Plate: CT808 row: M column: 2
Class: BAC ends
High quality sequence stop: 492.
FEATURES
Source .
location qualifiers
1..492
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=CT808 Col=2 Row=M"
/sex="M"
/clone_lib="CIT Human Genomic Sperm Library C"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"
ORIGIN
Query Match          93.6% Score 23.4; DB 9; Length 492;
Best Local Similarity 96.0%; Pred No.1.5e+03;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGCAACAAATCT 25
    |||||
Db 233 AAAAAAAAAATCGCAACAAATCT 209
    |||||

RESULT 4
LOCUS AG099199
DEFINITION Pan tregLdycies DNA, clone: PRB-101G12.F, genomic survey sequence.
ACCESSION AG099199
VERSION AG099199.1 GI:16719716
KEYWORDS GSS.
```

SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Pan troglodytes
REFERENCE AUTHORS	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarchontoglires; Primates; Catarrhini; Hominae; Pan.
TITLE	1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
JOURNAL	BAC end sequences of library PTB unpublished
REFERENCE AUTHORS	2 (bases 1 to 682) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
JOURNAL	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
COMMENT	PRIMERS Sequencing: -21M13
FEATURES	LIBRARY : PKS145 Vector : SacI R.Site 1 : SacI R.Site 2 : SacI. Location/Qualifiers 1..682 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" /clone="PTB-101G12.F" /sex="male" /cell_type="lymphoblast" /clone_id="PTB Chimpanzee Male BAC Library"
ORIGIN	Query Match 93.6%; Score 23.4; DB 10; Length 682; Best Local Similarity 96.0%; Pred. No. 1.5e+03; Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Dn	1 AAAAAAAAAATCGCAACAATCT 25 274 AAAAAAAAAATCGCAAAAATCT 298
RESULT 5	BH248096 795 bp DNA linear GSS 26-NOV-2001
LOCUS	BOGATS3TF BOGA Brassica oleracea genomic clone BOGATS3, genomic survey sequence.
DEFINITION	BH248096
ACCESSION	BH248096.1 GI:17068872
VERSION	GSS.
KEYWORDS	Brassica oleracea
SOURCE	Brassicaceae
ORGANISM	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Brassicales; Brassicaceae; Brassica.
REFERENCE	1 (bases 1 to 795) Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S., Utechtack,T.R., Wortman,J.R., White,O.R. and Town,C.D.
AUTHORS	Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis Genome Res. 15 (4), 487-495 (2005)
TITLE	Other GSSs: BOGATS3TR
JOURNAL	Contact: Chris Town
PUBMED	TIGR
COMMENT	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel.: 301-838-3523

Fax: 301-838-0208

Email: cdcowm@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: 7F

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..795
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T0100DH3"
/db_xref="taxon:3712"
/clone="BOGAT53"
/clone_1lb="BOGA"
/note="Vector: PHOS1; Site 1: BclXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BclXI linkers"

ORIGIN

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 9; Length 795;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy

1 AAAAAAAAAATCGCAACAATCT 25
255 AAAAAAAAAATCGCAACAATCT 231

Db

RESULT 6

BU509223/c 898 bp mRNA linear EST 12-SEP-2002

LOCUS AGENCOURT 10095473 NIH_MGC_71 Homo sapiens CDNA clone IMAGE:6503781

DEFINITION 5', mRNA sequence.

ACCESSION BU509223.1 GI:22815456

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.
1 (bases 1 to 898)

REFERENCE NIH-MGC http://mgc.ncl.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgsaps-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM14061 row: P column: 22

High quality sequence stop: 585.

FEATURES

source

Location/Qualifiers

1..898
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6503781"
/issue_type="telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

ORIGIN

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 5; Length 898;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25

Db

759 AAAAAAAAAATCGCAACAATCT 735

RESULT 7

LOCUS

AQ028539 296 bp DNA linear GSS 30-JUN-1998

DEFINITION CIT-HSP-2314C7.TR CIT-HSP Homo sapiens genomic clone 2314C7,

genomic survey sequence.

ACCESSION

AQ028539

VERSION

AQ028539.1 GI:3268761

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 296)
Adams,M.D., Rounseley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Simon,M., and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other GSSs: CIT-HSP-2314C7.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..296
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2314C7"
/sex="Male"
/cell_type="Sperm"
/clone_1lb="CIT-HSP"
/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2:
HindIII"

ORIGIN

Query Match

Best Local Similarity 88.0%; Score 22; DB 9; Length 296;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 AAAAAAAAAATCGCAACAATCT 22

96 AAAAAAAAAATCGCAACAATCT 75

Db

RESULT 8

BE154020/c 127 bp mRNA linear EST 21-JUN-2000

LOCUS PMO-HT0339-060400-009-F01 HT0339 Homo sapiens CDNA, mRNA sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 127)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=ct2=QV4-BT0257-261
400-009-F01&t3=2000-04-06&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 11

High quality sequence stop: 83.

FEATURES
Location/Qualifiers

source

1..127

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="HT0339"

/note="Organ: head,neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 127;
Best Local Similarity 92.0%; Pred. No. 5.4e+03;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25

Db 77 AAAAAAAAAATCGCAACAATCT 53

RESULT 9
BE062451

LOCUS QV4-BT0257-261099-011-b08 BT0257 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE062451 328 bp mRNA linear EST 09-JUN-2000

ACCESSION BE062451.1 GI:8407101

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 328)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,

O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=ct2=QV4-BT0257-261

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=ct2=QV4-BT0257-261
099-011-b08&t3=1999-10-26&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 7

High quality sequence stop: 328.

Location/Qualifiers

1..328

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="BT0257"

/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
Query Match 87.2%; Score 21.8; DB 2; Length 328;
Best Local Similarity 92.0%; Pred. No. 4.9e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25

Db 201 AAAAAAAAAATCGCAACAATCT 225

RESULT 10
BE062469/c

LOCUS QV4-BT0257-261099-011-g08 BT0257 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE062469 341 bp mRNA linear EST 09-JUN-2000

ACCESSION BE062469.1 GI:8407119

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 341)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,

O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=ct2=QV4-BT0257-261

099-011-g08&ct3=1999-10-26&ct4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 53
 High quality sequence stop: 341.
 Location/Qualifiers

FEATURES

source

1..341

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="BT0257"

/note="Organ: breast; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 341;
 Best Local Similarity 92.0%; Pred. No. 4.9e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAACCAATCT 25
 |||||
 155 AAAAAAAAAATCGCAACCAATCT 131

RESULT 11
 BF989405/c 375 bp mRNA linear EST 23-JAN-2001
 LOCUS CM3-GN0092-201000-407-g08 GN0092 Homo sapiens CDNA, mRNA sequence.

DEFINITION BF989405
 ACCESSION BF989405.1 GI:12395730
 VERSION EST.
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.

REFERENCE

1 (bases 1 to 375)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.,
 Nagai, M.A., de Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bai, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 PUBMED 10737800

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the PAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CM3&ct2=CM3-GN0092-
 201000-407-g08&ct3=2000-10-20&ct4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 8
 High quality sequence stop: 375.
 Location/Qualifiers

FEATURES

source

1..375

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"

/clone_lib="GN0092"
 /note="Organ: placenta_normal; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 375;
 Best Local Similarity 92.0%; Pred. No. 4.8e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAACCAATCT 25
 |||||
 347 AAAAAAAAAATCGCAACCAATCT 323

RESULT 12

AL947827 411 bp DNA linear GSS 02-APR-2004
 LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-307F02-015576,
 DEFINITION genomic survey sequence.

ACCESSION AL947827
 VERSION AL947827.1 GI:2440449
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosoid; eurosoid II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE

1 Li, Y., Rosso, M.G., Strizhov, N., Viehoever, P. and Weishaar, B.
 GABI-Kat Simplesearch: a flanking sequence tag (FST) database for
 the identification of T-DNA insertion mutants in Arabidopsis
 thaliana
 Bioinformatics 19 (11), 1441-1442 (2003)

REFERENCE

2 Rosso, M.G., Li, Y., Strizhov, N., Reis, B., Dekker, K. and
 Weishaar, B.
 An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for
 flanking sequence tag-based reverse genetics
 Plant Mol. Biol. 53 (1-2), 247-259 (2003)

REFERENCE

3 Strizhov, N., Li, Y., Rosso, M.G., Viehoever, P., Dekker, K.A. and
 Weishaar, B.

AUTHORS

High-throughput generation of sequence indexes from T-DNA
 mutagenized Arabidopsis thaliana lines
 Biotechniques 35 (6), 1164-1168 (2003)

TITLE

4 (bases 1 to 411)

Submitted (31-MAR-2004) Weishaar B., Max-Planck-Institut fuer
 Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 It indicates an insertion within the locus defined by BAC clone
 F15C21. Details on the protocols used for generation of the
 sequence are described in References 1-3. The sequences are
 generated at the MPI for Plant Breeding Research in the context of
 the GABI-Kat project. GABI-Kat is part of the German Plant Genomics
 program designated 'GABI'. Information on line availability can be
 found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.

COMMENT

Location/Qualifiers

FEATURES

source

1..411

/organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 /clone="GK-307F02-015576"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"

/ecotype="Col-0"
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161 (Genbank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed."

ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 411;
 Best Local Similarity 92.0%; Pred. No. 4.8e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
 |||||||
 DB 109 AAAAAAAAAATCGCAACAAATTT 133

RESULT 13
 BG461631 430 bp mRNA linear EST 21-APR-2001
 LOCUS RST44514 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG461631
 ACCESSION BG461631
 VERSION BG461631.1 GI:13750137
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 430)
 Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozar,S., Mays,R., Smith,B., Veloso,N., Klika,A., Hese,J., Cochren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.,
 Creation of genome-wide protein expression libraries using random activation of gene expression
 Nat. Biotechnol. 19 (5), 440-445 (2001)
 11329013
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@athersys.com.

FEATURES

source
 1..430
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_line="HT1080"
 /clone_lib="Athersys RAGE Library"
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 430;
 Best Local Similarity 92.0%; Pred. No. 4.8e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
 |||||||
 DB 293 AAAAAAAAAATCGCAACAAATCT 269

RESULT 14
 AQ117725/c

LOCUS AQ117725 449 bp DNA linear GSS 22-SEP-1998
 DEFINITION HS_2235_A1_F12_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2235 Col=23 Row=K, genomic survey sequence.

ACCESSION AQ117725
 VERSION AQ117725.1 GI:3495516
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 449)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

TITLE
 JOURNAL PUBLISHED
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2235 row: K column: 23
 Class: BAC ends
 High quality sequence stop: 449.

FEATURES
 source
 1..449
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=2235 Col=23 Row=K"
 /sex="male"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 449;
 Best Local Similarity 92.0%; Pred. No. 4.7e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
 |||||||
 DB 109 AAAAAAAAAATCGCAACAAATCT 85

RESULT 15
 AOS23511 538 bp DNA linear GSS 11-MAY-1999
 LOCUS HS_5197_A1_D04_TTA RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=773 Col=7 Row=G, genomic survey sequence.

ACCESSION AOS23511
 VERSION AOS23511.1 GI:4770908
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 538)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

TITLE
 JOURNAL PUBLISHED
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2235 row: K column: 23
 Class: BAC ends
 High quality sequence stop: 449.

COMMENT

Contact: Mahatras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 773 row: G column: 7
Seq primer: T7
Class: BAC ends
High quality sequence stop: 538.

FEATURES

Location/Qualifiers
1..538

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=773 Col=7 Row=G"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
BcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 538;
Best Local Similarity 92.0%; Pred. No. 4.7e+03;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25

Db 188 AAAAAAAAAATCGCAACAATCT 212

RESULT 16
CE254181

LOCUS tigr-gss-dog-17000336286588 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE254181
VERSION CE254181.1 GI:35953597
KEYWORDS GSS.

SOURCE
ORGANISM Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.

REFERENCE
AUTHORS 1 (bases 1 to 554)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.

TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627

COMMENT
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: shotgun.

FEATURES

Location/Qualifiers
1..554

/organism="Canis familiaris"
/mol_type="genomic DNA"

/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BclXI; Libraries were prepared from
peripheral blood"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 554;
Best Local Similarity 92.0%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25

Db 512 AAAAAAAAAATCGCAACAATCT 536

RESULT 17
BF187346

LOCUS BF187346 585 bp mRNA linear EST 10-MAR-2003
DEFINITION EST443633 potato stolon, Cornell University Solanum tuberosum cDNA
Clone CST440A6 5' sequence, mRNA sequence.
BF187346
BF187346.1 GI:11069565

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM Solanum tuberosum (potato)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamiales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 585)

REFERENCE
AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
Holt,I.E., Liang,F., Hansen,T.S., Utechtack,T., Bowman,C.L.,
Doen,B., Bougri,O., Buell,C.R., Konning,C.M., Tanksley,S.D. and
Baker,B.

Generation of ESTs from potato swelling stolons
Unpublished (1999)
Contact: Robin Buell

JOURNAL

9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org

This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/.

FEATURES

Location/Qualifiers

1..585

/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="CST440A6"
/tissue_type="axillary buds of stem explants, swelling
stolons"

/dev_stage="1 to 3 days"
/lab_host="SOLR"

/note="Vector: pluscript SK(-); Site 1: EcoRI; Site 2:
XhoI; RNA was supplied by Christian Bachem & Beatrix
Horvath(Laboratory of Plant Breeding, Dept. of Plant
Sciences, Wageningen University, The Netherlands). Total
RNA was isolated from developing axillary buds of potato
nodal stem cuttings cultured on medium for the
introduction of tuber formation as described in Bachem et
al. (Plant Journal 1996). Tissue samples were taken of
stages corresponding to growing stolons and the early
stages of tuber formation."

ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 585;
Best Local Similarity 92.0%; Pred. No. 4.6e+03;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25

|||||

Db 354 AAAAAAAAAATCAGAAAAAATCT 378

RESULT 18
LOCUS AQ18472
DEFINITION RPCI-11-177H1.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-177H1,
genomic survey sequence.
ACCESSION AQ18472
VERSION AQ18472.1 GI:4476196
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 604)
Zhaio,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: RPCI-11-177H1.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@igr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
Research Genet cs (<http://inforesgen.com>). BAC end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: S86
Classes: BAC ends.
FEATURES
source
1..604
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7567752"
/db_xref="taxon:9606"
/clone="RPCI-11-177H1"
/sex="Male"
/cell_type="Lymphocytes"
/clone_1fb="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"

ORIGIN
Query Match 87.2%; Score 21.8; DB 9; Length 604;
Best Local Similarity 92.0%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25
|||||
403 AAAAAAAAAATCGCAAAAAAACT 427

RESULT 19
LOCUS CV969982
DEFINITION 623 bp mRNA linear EST 25-JAN-2005
infestans CDNA, mRNA sequence.
ACCESSION CV969982
VERSION CV969982.1 GI:58159731
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

REFERENCE 1 (bases 1 to 623)
AUTHORS Randall,T., Dwyer,R.A., Huitema,E., Beyer,K., Cvitanich,C.,
Kelkar,H., Fong,A.M., Gates,K., Roberts,S., Yazkan,E., Gaffney,T.,
Law,M., Testa,A., Torto-Alalibo,A., Zhang,M., Zheng,L., Mueller,E.,
Windass,J., Binder,A., Birch,P.R.J., Giet,U., Govers,F., Gow,N.A.,
Mauch,F., van West,P., Maugh,M.E., Yu,J., Bolter,T., Kamoun,S.,
Lam,S.T. and Judelson, H.S.
TITLE Large-scale gene discovery in the oomycete Phytophthora infestans
reveals likely components of pathogenicity shared with true
fungi
JOURNAL Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)
PUBMED 15782637
COMMENT Contact: Judelson HS
Department of Plant Pathology
University of California
Webber Hall, Riverside, CA 92521, USA
Tel: 909 787 4199
Fax: 909 787 4294
Email: howard.judelson@ucr.edu
may be of host plant origin.
FEATURES
source
1..623
Location/Qualifiers
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="88069"
/db_xref="taxon:4787"
/sex="A1"
/clone_1fb="infected potato, center of lesion 6 dpi"
/note="Vector: pSPOR1"

ORIGIN
Query Match 87.2%; Score 21.8; DB 8; Length 623;
Best Local Similarity 92.0%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAATCT 25
|||||
371 AAAAAAAAAATCAGAAAAAATCT 395

Db 371 AAAAAAAAAATCAGAAAAAATCT 395

RESULT 20
LOCUS DE046629
DEFINITION Oryzias latipes DNA, clone: olal-012c20.R, genomic survey sequence.
ACCESSION DE046629
VERSION DE046629.1 GI:62558173
KEYWORDS GSS.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.
REFERENCE 1
Fujiyama,A., Toyoda,A., Kuroki,Y. and Sakaki,Y.
TITLE BAC end sequences of Olal Oryzias latipes library
JOURNAL Published Only in Database (2005)
AUTHORS 2 (bases 1 to 628)
Fujiyama,A.
TITLE Direct Submission
COMMENT Submitted (12-APR-2005) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:afujiyam@gsc.riken.jp, URL:<http://gsc.riken.jp/>,
Tel:81-3-4212-2558, Fax:81-3-3556-1916)
This work was done in collaboration with Takeda, H. (1), Naruse, K.
(2)
and Narita, T. (3)
(1) Department of Biological Science,
University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
Phone: +81-3-5841-4431
Fax: +81-3-5841-4993

```

E-mail: htakeda.s.u-tokyo.ac.jp
(2) Department of Biological Science,
University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
Phone: +81-3-5841-4431
Fax: +81-3-5841-4993
E-mail: naruse.s.u-tokyo.ac.jp
(3) Department of Biological Science,
University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
Phone: +81-3-5841-4431
Fax: +81-3-5841-4993
E-mail: tanarita.s.u-tokyo.ac.jp
PRIMERS
Sequencing : Forward
LIBRARY
Vector : pKS145
R.Site 1 : SacI
L.Site 2 : SacI.
Location/Qualifiers
1..628
/organism="Oryzias latipes"
/mol_type="genomic DNA"
/db_xref="taxon:8090"
/clone="olai-012C20.R"
/sex="male"
/cell_type="whole body"
/clone_lib="BAC end sequences of Olai Oryzias latipes
library"
ORIGIN
Query Match      87.2%; Score 21.8; DB 11; Length 628;
Best Local Similarity 92.0%; Pred. No. 4.5e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 AAAAAAAAAATCGCAACCAATCT 25
    |||||||
Db 490 AAAAAAAAAATCGCAACCAATCT 514

RESULT 21
CE354171 719 bp DNA linear GSS 26-SEP-2003
LOCUS tigr-gss-dog-17000361471950 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE354171 GI:36191598
VERSION CE354171.1 GI:36191598
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
          Buteyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Buteyola; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.
          1 (bases 1 to 719)
          Kirness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
          The dog genome: survey sequencing and comparative analysis
          Science 301 (5641), 1898-1903 (2003)
          14512627
          Contact: Kirness EF
          The Institute for Genomic Research
          Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
          Rockville, MD 20850, USA
          Tel: 301-838-0200
          Fax: 301-838-0208
          Email: ekirness@tigr.org
          Classes: shotgun.
FEATURES
source
location/Qualifiers
1..719
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"

```

```

/db_xref="taxon:9615"
/clone_lib="dog library"
/note="Site 1: Bact1; Libraries were prepared from
peripheral blood"
ORIGIN
Query Match      87.2%; Score 21.8; DB 10; Length 719;
Best Local Similarity 92.0%; Pred. No. 4.5e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 AAAAAAAAAATCGCAACCAATCT 25
    |||||||
Db 573 AAAAAAAAAATCGCAACCAATCT 549

RESULT 22
DR719016 764 bp mRNA linear EST 15-JUL-2005
LOCUS AGENCOURT_55091880 NIH_ZGC_7 Danio rerio cDNA clone IMAGE:791955
DEFINITION 5', mRNA sequence.
ACCESSION DR719016 GI:70893128
VERSION DR719016
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
          1 (bases 1 to 764)
          NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Daniela S. Gerhard, Ph.D.
          Office of Cancer Genomics
          National Cancer Institute / NIH
          Bldg. 31 Rm10A07 Bethesda, MD 20892
          Email: cgabs-remail.nih.gov
          Tissue Procurement: Ian Zon, Harvard
          cDNA Library Preparation: Open Biosystems
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LMNL at:
          http://image.llnl.gov
          Plate: LLM7141 row: h column: 01
          High quality sequence stop: 682.
FEATURES
source
location/Qualifiers
1..764
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:791955"
/tissue_type="whole body"
/lab_host="DH10B"
/clone_lib="NIH_ZGC_7"
/note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV.
Bulk tissue was collected from a whole adult individual
from the Tuebingen strain. 1st strand cDNA was primed with
a Not I - oligo(dT) primer, double-stranded cDNA was
cloned into the Not I and EcoRV sites of pExpress-1.
Library was size-selected for >1 kb fragments and
normalized. A non-normalized version of this library is
also available (NIH_ZGC_10). Library was constructed by
Open Biosystems (Huntsville, AL)"
ORIGIN
Query Match      87.2%; Score 21.8; DB 8; Length 764;
Best Local Similarity 92.0%; Pred. No. 4.5e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 AAAAAAAAAATCGCAACCAATCT 25
    |||||||
Db 39 AAAAAAAAAATCGCAACCAACT 15

```


RESULT 23
BH248808/c
LOCUS BH248808/c
DEFINITION BH248808 779 bp DNA linear GSS 26-NOV-2001
survey sequence.
ACCESSION BH248808
VERSION BH248808.1 GI:17070604
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 779)
Ayele, M., Haas, B.J., Kumar, N., Mu, H., Xiao, Y., Van Aken, S.,
Utterback, T.R., Mortman, J.R., White, O.R. and Town, C.D.,
Whole genome shotgun sequencing of Brassica oleracea and its
application to gene discovery and annotation in Arabidopsis
Genome Res. 15 (4), 487-495 (2005)
JOURNAL PUBMED
COMMENT Other GSSs: B0GAH70TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Classes: sheared ends.
Location/Qualifiers
1..779
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone="B0GAH70"
/clone_1lb="BOGA"
/note="Vector: PHOS1, Site 1: BstXI, 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN
Query Match 87.2%; Score 21.8; DB 9; Length 779;
Best Local Similarity 92.0%; Pred. No. 4.5e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
|||||
168 AAAAAAAAAATCGCAACAAACCT 144

RESULT 24
DN849376
LOCUS DN849376
DEFINITION 12888.2 Stolon Solanum tuberosum cDNA clone 12888 5', mRNA
sequence.
ACCESSION DN849376
VERSION DN849376.1 GI:62808536
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 796)
Flinn, B., Rothwell, C., Sardana, R., Griffiths, R., Lague, M., De
Koeyer, D., Andy, P., Goyer, C., Li, X.-Q., Wang-Pruski, G. and Regan, S.
Generation of ESTs from stolon tissues of potato
Unpublished (2004)
Contact: Barry Flinn
The Canadian Potato Genome Project - BioAtlantech

921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
Email: bflinn@bioatlantech.nb.ca
Seq primer: T3.
Location/Qualifiers
1..796
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Shepody"
/db_xref="taxon:4113"
/clone="12888"
/tissue_type="Stolon"
/lab_host="XLI0-Gold"
/clone_1lb="Stolon"
/note="Vector: pBluescript II SK(+) XR, Site 1: EcoRI;
Site 2: XhoI; supplier: Developmental series. Plants from
pathogen-free Solanum tuberosum var. Shepody, clone 1756,
nuclear stock were grown in a greenhouse under natural
conditions. RNA was isolated from stolon tissue."

ORIGIN
Query Match 87.2%; Score 21.8; DB 8; Length 796;
Best Local Similarity 92.0%; Pred. No. 4.5e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
|||||
317 AAAAAAAAAATCGCAACAAATCT 341

RESULT 25
BG890035
LOCUS BG890035
DEFINITION EST515886 cSTD Solanum tuberosum cDNA clone cSTD1604 5' sequence,
mRNA sequence.
ACCESSION BG890035
VERSION BG890035.1 GI:14267121
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 797)
van der Hoeven, R., Bezzerides, J., Ewing, E., Cho, J., Chiemiango, A.,
Bougril, O., Buell, C.R., Romling, C., Tankley, S. and Baker, B.
Generations of ESTs from dormant potato tubers
Unpublished (2001)
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M13F-R.
Location/Qualifiers
1..797
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD1604"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/clone_1lb="cSTD"
/note="Vector: pBluescript SK(-), Site 1: EcoRI, Site 2:
XhoI; This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tuber, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 40C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This

ORIGIN library is noted as p4 in Tankeley lab notebooks."

Query Match 87.2%; Score 21.8; DB 2; Length 797;
Best Local Similarity 92.0%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 294 AAAAAAAAAATCGCAACAAATCT 318

RESULT 26 BZ600741/c 820 bp DNA linear GSS 08-JUN-2003
LOCUS WHAB8741R Human MCF7 breast cancer cell line library (MCF7_1) Homo

DEFINITION sapiens genomic clone MCF7_1-706, genomic survey sequence.

ACCESSION BZ600741

VERSION BZ600741.1 GI:31509203

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 820)
Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q., Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.

End-sequence profiling: Sequence-based analysis of aberrant genomes Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)

12788976
Contact: Volik SV
Collin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com

Class: BAC ends.

FEATURES
source Location/Qualifiers
1..820
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-706"
/sex="female"
/clone_1ib="Human MCF7 breast cancer cell line library (MCF7_1)"
/note="Vector: pECBAC1; Site_1: HindIII; This library was constructed from MCF7 breast cancer cell line by Amplicon Express (http://www.genomex.com) using their standard procedure."

ORIGIN
Query Match 87.2%; Score 21.8; DB 9; Length 820;
Best Local Similarity 92.0%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 114 AAAAAAAAAATCGCAACAAATCT 90

RESULT 27 BZ605294 834 bp DNA linear GSS 08-JUN-2003
LOCUS WHAB0261R Human MCF7 breast cancer cell line library (MCF7_1) Homo
DEFINITION sapiens genomic clone MCF7_1-10F3, genomic survey sequence.
ACCESSION BZ605294
VERSION BZ605294.1 GI:31513855

KEYWORDS GSS.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 834)
Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q., Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.

End-sequence profiling: Sequence-based analysis of aberrant genomes Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)

12788976
Contact: Volik SV
Collin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com

Class: BAC ends.

FEATURES
source Location/Qualifiers
1..834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-10F3"
/sex="female"
/clone_1ib="Human MCF7 breast cancer cell line library (MCF7_1)"
/note="Vector: pECBAC1; Site_1: HindIII; This library was constructed from MCF7 breast cancer cell line by Amplicon Express (http://www.genomex.com) using their standard procedure."

ORIGIN
Query Match 87.2%; Score 21.8; DB 9; Length 834;
Best Local Similarity 92.0%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 482 AAAAAAAAAATCGCAACAAATCT 506

RESULT 28 BZ109269/c 846 bp DNA linear GSS 11-OCT-2002
LOCUS CH230-135H18.TUB CHORI-230 Segment 1 Rattus norvegicus genomic
DEFINITION clone CH230-135H18, genomic survey sequence.
ACCESSION BZ109269
VERSION BZ109269.1 GI:23750148

KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 846)
Zhao,S., Shetty,J., Shatman,S., Teagay,G., Geer,K., Shvartsbeyn,A., Gebregorgis,B., Overton,L., Russell,D., Chen,D., Rigns,F., de Jong,P. and Frazer,C.M.

Rat BAC End sequences from Library CHORI-230 EcORI segment Unpublished (1999)

CONTACT: Shaving Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(<http://www.chori.org/bacpac/oreringinformation.html>). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 135 row: H column: 18

Seq primer: SP6
Classes: BAC ends.

FEATURES

source

1. 846
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNhsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-135H18"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/note="Vector: PTABAC2.1; Site_1: EcoRI, Site_2: EcoRI;
CHORI-230 Rat (BN/SsNhsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 846;
Best Local Similarity 92.0%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 33 AAAAAAAAAATCGCAACAAATCT 9

RESULT 29 A0745442 875 bp DNA linear GSS 16-JUL-1999
LOCUS A0745442/c HS_2278_A1_B12_T7C CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2278 Col=23 Row=C, genomic survey
sequence.

ACCESSION A0745442
VERSION A0745442.1 GI:5522964
KEYWORDS GSS.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 875)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

TITLE Contact: Mahairas GG, Wallace JC, Hood L
JOURNAL High Throughput Sequencing Center
PUBMED 10499764
COMMENT 401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 2278 row: C column: 23
Seq primer: T7
Classes: BAC ends
High quality sequence stop: 875.

FEATURES
source
1. 875
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

/clone="Plate=2278 Col=23 Row=C"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH108"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 875;
Best Local Similarity 92.0%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 169 AAAAAAAAAATCGCAACAAATCT 145

RESULT 30 DU022055/c 888 bp DNA linear GSS 12-AUG-2005
LOCUS DU022055/c 1790 Tomato HindIII BAC Library Lycopersicon esculentum genomic
DEFINITION clone LE_HBa0005F08 3, genomic survey sequence.
ACCESSION DU022055
VERSION DU022055.1 GI:72426062
KEYWORDS GSS.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 888)
Mueller,L.A., Buelis,R.M., Wang,Y., Tankeley,S.D., Giovannoni,J.J.,
Van Eck,J. and Stack,S.
BAC end sequencing from three Solanum lycopersicon libraries
Other GSSs: 1791

TITLE Contact: Lukas Mueller
JOURNAL Tanksley Lab, Dept. of Plant Breeding
PUBMED Cornell University
COMMENT 251 Emerson Hall, Ithaca, NY 14853, USA
Tel: 607-255-6557
Fax: 607-255-6683
Email: sgn-feedback@sgn.cornell.edu
Insert Length: 87902 Std Error: 0.00
Plate: 5 row: F column: 8
Seq primer: SP6
Classes: BAC ends
High quality sequence start: 16
High quality sequence stop: 796.

FEATURES
source
1. 888
Location/Qualifiers
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/cultivar="Heinz 1706"
/db_xref="taxon:4081"
/clone="LE_HBa0005F08"
/lab_host="E. coli"
/clone_lib="Tomato HindIII BAC Library"
/note="Vector: pBelOBAC11; Site_1: HindIII"

ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 888;
Best Local Similarity 92.0%; Pred. No. 4.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAATCT 25
Db 838 AAAAAAAAAATCGCAACAAATCT 814

RESULT 31 CC212315 1163 bp DNA linear GSS 12-MAY-2003
LOCUS CC212315 CH261-23A11_Sp6.1 CH261 Gallus gallus genomic clone CH261-23A11,
DEFINITION

genomic survey sequence.
ACCESSION CC212315
VERSION CC212315.1 GI:30530983
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1163)
AUTHORS Krentliki, C., Higgsbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E., and Wilson, R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert length: 18200 Std Error: 0.00
Seq primer: SP6 ATTTAGTGACACTATAG
Class: BAC ends
High quality sequence start: 57
High quality sequence stop: 755.
Location/Qualifiers
1..1163
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-23A11"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_1fb="CH261"
/note="Vector: pTRABAC2.1; Site 1: EcoRI, Site 2: EcoRI;
CH261 Female Chicken library - for library and clone
ordering information: <http://www.chori.org/bacpac>"

ORIGIN
Query Match 87.2%; Score 21.8; DB 9; Length 1163;
Best Local Similarity 92.0%; Pred. No. 4.3e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 |AAAAAAAAAATCGCAACAATCT 25
DB 187 AAAAAAAAAATCACAACAATAT 211

RESULT 32
AJ927038 1437 bp mRNA linear EST 10-JUN-2005
LOCUS AJ927038 Theileria annulata merozoite Theileria annulata cDNA clone
DEFINITION tam023g05.plk, mRNA sequence.
ACCESSION AJ927038
VERSION AJ927038.1 GI:67497421
KEYWORDS EST.
SOURCE Theileria annulata
ORGANISM Theileria annulata
Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
Theileria.
REFERENCE 1 (bases 1 to 1437)
AUTHORS Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C. A., Weir, W.,
Kerhournou, A., Aslett, M., Bishop, R., Bouchier, C., Cochet, M.,
Coulson, R. M., Cronin, A., de Villiers, E., Fraser, A., Foster, N.,
Gardner, M., Goble, A., Griffiths-Jones, S., Harris, D. E., Katzer, F.,
Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F.,
Nene, V., O'Neill, S., Price, C., Quail, M. A., Rabinowitsch, E.,
Rawlings, N. D., Rutter, S., Saunders, D., Seeger, K., Shah, T.,
Squares, R., Squares, S., Tivey, A., Walker, A. R., Woodward, J.,
Dobeleare, D. A. E., Langley, G., Rajandream, M. A., McKeever, D.,
Shiels, B., Tait, A., Barrell, B., and Hall, N.
TITLE The genome of the host-cell transforming parasite Theileria
annulata and a comparison with T. parva
JOURNAL Unpublished (2005)

COMMENT Contact: Pain A
The Pathogen Sequencing Unit
The Wellcome Trust Sanger Institute
Genome Campus, CB10 1SA, UNITED KINGDOM
Merozoite cDNA library: Frank Katzer and Brian Shiels, Division of
Veterinary Infection and Immunity, ICM, University of Glasgow, UK.
FEATURES
source
Location/Qualifiers
1..1437
/organism="Theileria annulata"
/mol_type="mRNA"
/isolate="Ankara (clone D7)"
/db_xref="taxon:5874"
/clone="tam023g05.plk"
/dev_stage="merozoite"
/lab_host="Bos taurus (cow)"
/clone_1fb="Theileria annulata merozoite"
/note="country: Turkey:Ankara"

ORIGIN
Query Match 87.2%; Score 21.8; DB 1; Length 1437;
Best Local Similarity 92.0%; Pred. No. 4.2e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 |AAAAAAAAAATCGCAACAATCT 25
DB 33 AAAAAAAAAATCTCAAAAAAAAAATCT 9

RESULT 33
CO873113 397 bp mRNA linear EST 01-SEP-2004
LOCUS CO873113
DEFINITION R2PDP105601752Q 5', mRNA sequence.
ACCESSION R2PDP105601752Q 5', mRNA sequence.
VERSION CO873113
KEYWORDS CO873113.1 GI:51802953
SOURCE EST.
ORGANISM Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 397)
AUTHORS Hennig, S., Jantitz, M., Herwig, R., and Williams, J.
TITLE Generation, annotation, evolutionary analysis and database
integration of 14969 cattle EST clusters
JOURNAL Unpublished (2004)
COMMENT Contact: Hennig S
Laboratory 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones that display the same hybridisation matrix with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per ONFP cluster was selected for sequencing. cDNA clones and
filters are distributed via Deutsches Ressourcenzentrum fuer
Genomforschung GmbH (<http://www.rzpd.de>).
PCR primers
FORWARD: 5' CCCGAGGCTTACACTTATACCTCCGCTCG 3' (M13RSP) 5'-seg
BACKWARD: 5' GCATATCGCGACTCGCGAAGGGGATGTG 3' (M13FSP) 3'-seg
Seq primer: 5'-CCGGTCGGAATTCGCCGGGT-3' (M13RSP).
Location/Qualifiers
1..397
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="R2PDP105601752Q"
/sex="female"
/tissue_type="brain tissue"
/dev_stage="adult brain"

/clone.lib="normal cattle brain"
/note="Organ: brain; Vector: pSport1; Site 1: NotI;
Site 2: SalI; Random primed and directionally cloned in
pSport1 vector using NotI
(5'-pGACTGATCTAGATCGAGCGCGCCGCC (P)15-3' and SalI 5'-
TCGACCGACGCGTCCG-3' adapters (Gibco BRL))"

Best Local Similarity 88.0%; Pred. No. 6.4e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACAATCT 25
Db 341 AAAAAAAAAANNCACAAATAT 317

Query Match 85.6%; Score 21.4; DB 7; Length 397;
Best Local Similarity 88.0%; Pred. No. 6.4e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACAATCT 25
Db 310 AAAAAAAAAANNCACAAATAT 334

RESULT 35
LOCUS CF972324 451 bp mRNA linear EST 24-NOV-2003
DEFINITION VRJ9 vitis riparia endodermant bud - VRJ vitis riparia cDNA clone
VRJ9 3', mRNA sequence.
ACCESSION CF972324
VERSION CF972324.1 GI:38498394
KEYWORDS EST.
SOURCE Vitis riparia
ORGANISM Vitis riparia
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; Vitaceae; Vitis.
1 (bases 1 to 451)
Fennell, A. and Mathiason, K.
Expressed sequence tags from endodermant Vitis riparia buds
Unpublished (2003)
Contact: Anne Fennell
Horticulture, Forestry, Landscape and Parks Department
South Dakota State University
Northern Plains Bioresearch Laboratory, Box 2140A, Brookings, SD
57007-0996, USA
Tel: 605 688 6373
Fax: 605 688 4713
Email: Anne.Fennell@state.edu
Seq primer: SP6.

REFERENCE 1 (bases 1 to 420)
AUTHORS Rogatcheva, M.B., Meyers, S., He, W., Larkin, D.M., Marron, B.M.,
Beever, J.E. and Schook, L.B.
TITLE Piggy-BACing the Human Genome: Constructing a Porcine Physical Map
Through Comparative Genomics
Unpublished (2004)
Other GSSs: RPCI44_443L23.f
Contact: Lawrence B. Schook
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 265 5326
Fax: 217 244 5617
Email: schook@uiuc.edu
Clones are derived from the porcine BAC library RPCI-44
(http://www.bacpac.chori.org/porcine242.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@chori.org).
Clones may be purchased from BACPAC Resources
(http://BACPACresources.chori.org). This work was undertaken as part
of the International Swine Genome Sequencing Consortium by
University of Illinois at Urbana Champaign, USA with funds provided
by grant No. AG2002-34480-11828 from USDA-CGRS and
AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
Initiative)
Plate: 443 row: L column: 23
Seq primer: SP6
Classes: BAC ends.

FEATURES
source 1.451
Location/Qualifiers
/organism="Vitis riparia"
/mol_type="mRNA"
/db_xref="taxon:96939"
/clone="VRJ9"
/dev_stage="dormant"
/lab_host="DH10B"
/clone_lib="Vitis riparia endodermant bud - VRJ"
/note="Organ: bud; Vector: pSport 1; Site 1: SalI; Site 2:
NotI; VRJ is a cDNA library of Vitis riparia endodermant
buds. Endodermant buds were collected from insect and
disease free vines that were induced into dormancy with
short photoperiods. The directionally oriented library was
constructed according to Gibco BRL SuperScript Plasmid
System for cDNA synthesis and plasmid cloning."

Journal
COMMENT
Other GSSs: RPCI44_443L23.f
Contact: Lawrence B. Schook
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 265 5326
Fax: 217 244 5617
Email: schook@uiuc.edu
Clones are derived from the porcine BAC library RPCI-44
(http://www.bacpac.chori.org/porcine242.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@chori.org).
Clones may be purchased from BACPAC Resources
(http://BACPACresources.chori.org). This work was undertaken as part
of the International Swine Genome Sequencing Consortium by
University of Illinois at Urbana Champaign, USA with funds provided
by grant No. AG2002-34480-11828 from USDA-CGRS and
AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
Initiative)
Plate: 443 row: L column: 23
Seq primer: SP6
Classes: BAC ends.

FEATURES
source 1.420
Location/Qualifiers
/organism="Sus scrofa"
/mol_type="genomic DNA"
/strain="four pigs (bred: 37.5% Yorks Landrace and 25%
Meishan)"
/db_xref="taxon:9823"
/clone="RPCI44_443L23"
/sex="male"
/cell_type="blood"
/clone_lib="RPCI-44"
/note="Vector: pTARBAC2; Site 1: EcoRI; Site 2: EcoRI;
porcine male BAC library produced by Pieter de Jong"

Query Match 85.6%; Score 21.4; DB 7; Length 451;
Best Local Similarity 95.7%; Pred. No. 6.3e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACAAT 23
Db 114 AAAAAAAAAATCGCAACAAT 136

RESULT 36
LOCUS BF054075 550 bp mRNA linear EST 07-MAR-2003
DEFINITION CSTB3812 5' sequence, mRNA sequence.
ACCESSION BF054075
VERSION BF054075.1 GI:10807971
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

Query Match 85.6%; Score 21.4; DB 10; Length 420;
Best Local Similarity 95.7%; Pred. No. 6.3e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGCAACAAT 23
Db 114 AAAAAAAAAATCGCAACAAT 136

ACCESSION CV434921
 VERSION CV434921.1 GI:52844211
 KEYWORDS EST.
 SOURCE Solanum tuberosum (potato)
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanales; Solanales; Solanum.
 1 (bases 1 to 770)
 REFERENCE Plim,B., Rothwell,C., Sardana,R., Griffiths,R., Laque,M., De Koeper,D., Audy,P., Goyer,C., Li,X.-O., Wang-Pruski,G. and Regan,S. Generation of ESTs from potato suspension cultures
 TITLE Unpublished (2004)
 JOURNAL Contact: Barry Plim
 COMMENT The Canadian Potato Genome Project - BioAtlantech
 921 Canadian Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
 Email: bplim@biatlantech.nb.ca
 Seq primer: T3
 FEATURES
 source Location/Qualifiers
 1..770
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Shepody"
 /db_xref="taxon:4113"
 /clone="58424"
 /tissue_type="Callus-derived suspension culture"
 /lab_host="X110-Gold"
 /clone_lib="Suspension culture"
 /note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; supplier: Developmental series. Callus was induced from Shepody, clone 1736, sterile stem sections by culture on Callus Induction Medium (CIM), comprised of MS medium (pH 5.6) containing 10 mg/L Thiamine-HCL, .01 mg/L Kinetin and 3 mg/L 2,4-D solidified with .8% (w/v) Phytagar. Suspensions are induced by placing callus from the plates into 125 ml Erlenmeyer flasks with liquid CIM (no Phytagar) at a density of 10% (w/v) in volumes of approximately 30-35 ml. Cells were subcultured weekly by transfer to fresh media, with the density remaining at 10% (w/v) and the volume remaining around 30 ml. Cells were collected for RNA isolations and library construction 5 days after subculture."

ORIGIN
 Query Match 85.6%; Score 21.4; DB 7; Length 770;
 Best Local Similarity 95.7%; Pred. No. 6e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAAAT 23
 |||||||||
 119 AAAAAAAAAATCGCAACAAAT 97
 RESULT 40
 DU036823 841 bp DNA linear GSS 12-AUG-2005
 LOCUS 19481 Tomato HindIII BAC library Lycopersicon esculentum genomic
 ACCESSION clone LE_HBA0078N21 5, genomic survey sequence.
 VERSION DU036823
 KEYWORDS GSS.
 SOURCE Lycopersicon esculentum (Solanum lycopersicum)
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanales; Solanum; Lycopersicon.
 1 (bases 1 to 841)
 REFERENCE Mueller,L.A., Bueis,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J., Van Eck,J.L. and Stack,S. BAC end sequencing from three Solanum lycopersicon libraries
 TITLE Unpublished (2005)
 JOURNAL Other GSSs: 19480
 COMMENT Contact: Lukas Mueller

Tanksley Lab, Dept. of Plant Breeding
 Cornell University
 251 Emerson Hall, Ithaca, NY 14853, USA
 Tel: 607-255-6557
 Fax: 607-255-6683
 Email: sgn-feedback@cornell.edu
 Plate: 78 row: N column: 21
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 49
 High quality sequence stop: 841.
 FEATURES
 source Location/Qualifiers
 1..841
 /organism="Lycopersicon esculentum"
 /mol_type="genomic DNA"
 /cultivar="Heinz 1706"
 /db_xref="taxon:4081"
 /clone="LE_HBA0078N21"
 /lab_host="E. coli"
 /clone_lib="Tomato HindIII BAC library"
 /note="Vector: pBlotBAC11; Site 1: HindIII"

ORIGIN
 Query Match 85.6%; Score 21.4; DB 10; Length 841;
 Best Local Similarity 95.7%; Pred. No. 5.9e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGCAACAAAT 23
 |||||||||
 244 AAAAAAAAAATCGCAACAAAT 266
 RESULT 41
 CDS1FMG 999 bp DNA linear GSS 01-JUN-2001
 LOCUS Anopheles gambiae GSS SP6 end of clone 04M19 of NotreDame1 library
 DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.
 ACCESSION AL142025
 VERSION AL142025.1 GI:7000143
 KEYWORDS GSS.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Anophelinae; Anopheles.
 1 (bases 1 to 999)
 REFERENCE Genoscope.
 DIRECT SUBMISSION Direct Submission
 SUBMITTED (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segrafe@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 2 (bases 1 to 999)
 REFERENCE Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weisembach,J. Submitted (16-FEB-2000) BIMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
 This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
 FEATURES
 source Location/Qualifiers
 1..999
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="P87"
 /db_xref="taxon:7165"
 /clone="04M19"
 /clone_lib="NotreDame1"
 /note="end : SP6"

ORIGIN
 Query Match 85.6%; Score 21.4; DB 10; Length 999;

Best Local Similarity 95.7%; Pred. No. 5.8e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAAT 23
Db 582 AAAAAAAAAATCGCAACCAAT 604

RESULT 42
LOCUS CL051409/c 1043 bp DNA linear GSS 31-DEC-2003

DEFINITION CH216-73N12.RM1.1 CH216 Xenopus tropicalis genomic clone
CH216-73N12, genomic survey sequence.

ACCESSION CL051409
VERSION CL051409.1 GI:40507322

KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

REFERENCE 1 (bases 1 to 1043)
AUTHORS Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson

Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM1 TACGACTCCTATACGAGAGA
Class: BAC ends

High quality sequence start: 8
High quality sequence stop: 714.
Location/Qualifiers
1..1043

FEATURES
source
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-73N12"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHOR1-216 Xenopus tropicalis
BAC library"

ORIGIN
Query Match 85.6%; Score 21.4; DB 10; Length 1043;
Best Local Similarity 95.7%; Pred. No. 5.8e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAAT 23
Db 385 AAAAAAAAAATCGCAACCAAT 363

RESULT 43
LOCUS CN240069 1991 bp mRNA linear EST 09-APR-2004
DEFINITION EST005928 Mycelium and Yeast cells from Paracoccidioides
brasiliensis Paracoccidioides brasiliensis cDNA, mRNA sequence.

ACCESSION CN240069
VERSION CN240069.1 GI:46343813
KEYWORDS EST.
SOURCE Paracoccidioides brasiliensis
Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Ouygenales; Mitosporic Ouygenales; Paracoccidioides.

REFERENCE 1 (bases 1 to 1991)
AUTHORS Felipe, M.S., Andrade, R.V., Arraes, F.B., Nicola, A.M., Maranhao, A.O.,
Torres, F.A., Silva-Pereira, I., Pocas-Fonseca, M.J., Campos, E.G.,

Moraes, L.M., Andrade, P.A., Tavares, A.H., Silva, S.S., Kyaw, C.M.,
Souza, D.P., Network, P., Pereira, M., Jesuino, R.S., Andrade, E.V.,
Pachine, J.A., Oliveira, G.S., Barbosa, M.S., Martins, N.F.,
Fachin, A.L., Cardoso, R.S., Passos, G.A., Almeida, N.F., Walter, M.E.,
Soares, C.M., Carvalho, M.J. and Brigido, M.M.
Transcriptional Profiles of the Human Pathogenic Fungus
Paracoccidioides brasiliensis in Mycelium and Yeast Cells
J. Biol. Chem. 280 (26), 24706-24714 (2005)
15849188
CONTACT: Felipe MSS
LABORATORY OF MOLECULAR BIOLOGY
INSTITUTE OF BIOLOGY - UNIVERSITY OF BRASILIA
CAMPUS UNIVERSITARIO, ASA NORTE, BRASILIA, DF 70910-900, BRA
Tel: 55 61 307 2423
Fax: 55 61 349 8411
Email: msuelli@unb.br

FEATURES
source
Location/Qualifiers
1..1991

/organism="Paracoccidioides brasiliensis"
/mol_type="mRNA"
/strain="PD01"
/db_xref="taxon:121759"
/clone_lib="Mycelium and yeast cells from Paracoccidioides
brasiliensis"
/note="Pb Lambda Zap Express library"

ORIGIN
Query Match 85.6%; Score 21.4; DB 7; Length 1991;
Best Local Similarity 95.7%; Pred. No. 5.4e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACCAAT 23
Db 1082 AAAAAAAAAATCGCAACCAAT 1104

RESULT 44
LOCUS AK039894/c 2078 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
library, clone:A430025117 product:unknown EST, full insert
sequence.
ACCESSION AK039894
VERSION AK039894.1 GI:26087475
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
PUBMED 10349636

REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, Y., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M.,
Sunt, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Matsubara, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multiplexed sequencer

FEATURES	source
COMMENT	<p>CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.</p> <p>Please visit our web site for further details.</p> <p>URL: http://genome.gsc.riken.jp/</p> <p>URL: http://fantom.gsc.riken.jp/</p>
FEATURES	<p>Location/Qualifiers</p> <p>1..2078</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="CS78L/6J7"</p> <p>/db_xref="PANTOM,DB:A430025117"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="A430025117"</p> <p>/cissue_type="chymus"</p> <p>/clone_lib="RIKEN full-length enriched mouse cDNA library"</p> <p>/dev_stage="0 day neonate"</p> <p>1..2078</p> <p>/note="unknown EST [GB BB201599, evidence: BLASTN, 96%, match=220]"</p>
ORIGIN	<p>misc_feature</p> <p>1..2078</p> <p>/note="unknown EST [GB BB201599, evidence: BLASTN, 96%, match=220]"</p>
Query Match	85.6%; Score 21.4; DB 4; Length 2078;
Best Local Similarity	95.7%; Freq. No. 5.4e+03;
Matches	22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	1185 AAAAAAAAAATCGCAACCAAT 1163
Qy	1 AAAAAAAAAATCGCAACCAAT 23
LOCUS	AU133371 774 bp mRNA linear EST 01-AUG-2002
DEFINITION	AU133371 NT2RNP4 Homo sapiens cDNA clone NT2RNP4001940 5', mRNA
ACCESSION	AU133371

VERSION AU133371.1 GI:10993910
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 774)
 AUTHORS Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J.,
 Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuko,Y. and
 Iisogai,T.
 TITLE HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y.,
 Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T.,
 Sugano,S., Masuko,Y., Iisogai,T.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Iisogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomcshri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 FEATURES
 source
 1..774
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RP4001940"
 /cell_type="teratocarcinoma"
 /cell_line="NT2"
 /clone_id="NT2RP4"
 /note="Vector: PME18SPJ3; mRNA from NT2 neuronal precursor
 cells after 2-weeks retinoic acid (RA) induction"
 ORIGIN
 Query Match 84.0%; Score 21; DB 1; Length 774;
 Best Local Similarity 95.5%; Prod. No. 7,9e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAATCGCAACAAA 22
 ||||||||||||||||||||
 Db 699 AAAAAAAAAATCGCAACAAA 678
 RESULT 46
 CDS00FLB 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
 DEFINITION BACR3C10 of RPE1-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 AL070935
 AL070935.1 GI:4950775
 GSS.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 REFERENCE Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FR
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila


```

VERSION      CK609736.1  GI:41330794
KEYWORDS
SOURCE       Gallus gallus (chicken)
ORGANISM     Gallus gallus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
              Phasianinae; Gallus.
REFERENCE    1 (bases 1 to 203)
AUTHORS     Bliss,T.W., Emara,M.G., Dohms,J.E., Lillehoj,H.S. and Keeler,C.L.
TITLE       Analysis of gene expression in differentially stimulated avian
              macrophages
JOURNAL      Unpublished (2004)
COMMENT      Contact: Calvin Keeler
              Dept. of Animal and Food Sciences
              University of Delaware
              040 Townsend Hall, Newark, DE 19716-2150, USA
              Tel: 302-831-6473
              Fax: 302-831-2822
              Email: ckeeler@udel.edu
              www.aviangenomics.udel.edu
PCR Primers  FORWARD: T3
              BACKWARD: T7.
FEATURES
  source      Location/Qualifiers
              1..203
                /organism="Gallus gallus"
                /mol_type="mRNA"
                /strain="Broiler"
                /db_xref="taxon:9031"
                /clone="IFNK_L12"
                /sex="Female and male combined"
                /tissue_type="Blood"
                /cell_type="Macrophage"
                /clone_1fb="Interferon Stimulated Chicken PBL Macrophage"
                /note="Vector: pbluescript SK+; Site_1: EcoRI, Site_2:
                XhoI"

ORIGIN
Query Match      83.2%; Score 20.8; DB 7; Length 203;
Best Local Similarity 91.7%; Pred. No. 1.e+04;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATC 24
    |||||
Db 138 AAAAAAAAAATCGCAACCAATC 115

RESULT 50
AI680501      225 bp      mRNA      linear      EST 15-DEC-1999
LOCUS         tw63f08.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2264391 3',
DEFINITION   mRNA sequence.
ACCESSION    AI680501
VERSION      AI680501.1  GI:4890683
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homidae; Homo.
REFERENCE    1 (bases 1 to 225)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D.
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/ILNL at:
www.bio.lnlnl.gov/bbrp/image/image.html
Insert Length: 1478 Std Error: 0.00
Seq Primer: -40UP from Gibco
High quality sequence stop: 219.
FEATURES
  source      Location/Qualifiers
              1..225
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:2264391"
                /tissue_type="poorly-differentiated endometrial
                adenocarcinoma, 2 pooled tumors"
                /lab_host="DH10B"
                /clone_1fb="NCI_CGAP_Ut3"
                /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 1.45 kb. Life Technologies catalog #:
                11541-018"

ORIGIN
Query Match      83.2%; Score 20.8; DB 1; Length 225;
Best Local Similarity 91.7%; Pred. No. 1.e+04;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACCAATC 24
    |||||
Db 160 AAAAAAAAAATCGCAACCAATC 183

Search completed: December 14, 2005, 07:34:29
Job time : 1764.1 secs

```



```
C 97 18.2 72.8 598 6 US-10-750-185-4608
C 98 18.2 72.8 699 6 US-10-750-185-28395
C 99 18.2 72.8 1051 6 US-10-750-185-38723
C 100 18.2 72.8 1135 6 US-10-750-185-60012
C 101 18.2 72.8 1340 6 US-10-750-185-25554
C 102 18.2 72.8 1400 6 US-10-750-185-46908
C 103 18.2 72.8 1476 6 US-10-750-185-37578
C 104 18.2 72.8 1555 6 US-10-750-185-50236
C 105 18.2 72.8 1709 6 US-10-750-185-27058
C 106 18.2 72.8 2141 6 US-10-750-185-43265
C 107 18.2 72.8 3128 6 US-10-821-234-831
C 108 18.2 72.8 3225 6 US-10-750-185-53852
C 109 18.2 72.8 4065 6 US-10-750-185-51154
C 110 18.2 72.8 124972 7 US-11-121-086-100
C 111 18.2 72.8 134174 7 US-11-121-086-99
C 112 18.2 72.8 137671 7 US-11-121-086-47
C 113 18.2 72.8 141121 6 US-10-995-561-13262
C 114 18.2 72.8 167116 7 US-11-121-086-44
C 115 18.2 72.8 168516 7 US-11-121-086-3
C 116 18.2 72.8 180654 7 US-11-121-086-58
C 117 18.2 72.8 195235 6 US-10-995-561-13495
C 118 18.2 72.8 215308 7 US-11-121-086-77
C 119 18.2 72.8 380749 6 US-10-995-561-13216
C 120 17.8 71.2 201 6 US-10-750-185-2006
C 121 17.8 71.2 600 6 US-10-750-185-52410
C 122 17.8 71.2 1066 6 US-10-750-185-40600
C 123 17.8 71.2 1497 6 US-10-750-185-64000
C 124 17.8 71.2 1704 6 US-10-750-185-63557
C 125 17.8 71.2 1758 6 US-10-750-185-41807
C 126 17.8 71.2 2409 6 US-10-689-742-73
C 127 17.8 71.2 2503 6 US-10-750-185-38626
C 128 17.8 71.2 3454 6 US-10-750-185-32866
C 129 17.8 71.2 49979 6 US-11-121-086-71
C 130 17.8 71.2 164527 7 US-11-121-086-89
C 131 17.8 71.2 201990 6 US-10-995-561-13303
C 132 17.8 71.2 201990 6 US-10-995-561-18388
C 133 17.8 71.2 201990 6 US-10-995-561-42070
C 134 17.6 70.4 201 6 US-10-995-561-42571
C 135 17.6 70.4 201 6 US-10-995-561-42572
C 136 17.6 70.4 201 6 US-10-995-561-44300
C 137 17.6 70.4 201 6 US-10-995-561-44302
C 138 17.6 70.4 201 6 US-10-995-561-54933
C 139 17.6 70.4 201 6 US-10-995-561-62530
C 140 17.6 70.4 201 6 US-10-995-561-62532
C 141 17.6 70.4 201 6 US-10-995-561-62532
C 142 17.6 70.4 201 6 US-10-995-561-67774
C 143 17.6 70.4 201 6 US-10-995-561-76811
C 144 17.6 70.4 201 6 US-10-995-561-76977
C 145 17.6 70.4 201 6 US-10-995-561-76978
C 146 17.6 70.4 482 6 US-10-955-054A-36
C 147 17.6 70.4 600 6 US-10-750-185-839
C 148 17.6 70.4 600 6 US-10-750-185-2046
C 149 17.6 70.4 600 6 US-10-750-185-3096
C 150 17.6 70.4 881 6 US-10-750-185-33928
```

ALIGNMENTS

```
RESULT 1
US-11-121-086-102
; Sequence 102, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138, 6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
```

```
Sequence 4608, Ap
Sequence 28395, A
Sequence 38723, A
Sequence 60012, A
Sequence 25554, A
Sequence 46908, A
Sequence 37578, A
Sequence 50236, A
Sequence 27058, A
Sequence 43265, A
Sequence 831, App
Sequence 53852, A
Sequence 51154, A
Sequence 100, App
Sequence 99, Appl
Sequence 47, Appl
Sequence 13262, A
Sequence 44, Appl
Sequence 3, Appl
Sequence 58, Appl
Sequence 13495, A
Sequence 77, Appl
Sequence 13216, A
Sequence 49975, A
Sequence 2006, Ap
Sequence 52410, A
Sequence 40600, A
Sequence 64000, A
Sequence 63557, A
Sequence 41807, A
Sequence 73, Appl
Sequence 38626, A
Sequence 32866, A
Sequence 13443, A
Sequence 71, Appl
Sequence 89, Appl
Sequence 13303, A
Sequence 18388, A
Sequence 42070, A
Sequence 42571, A
Sequence 42572, A
Sequence 44300, A
Sequence 44302, A
Sequence 54933, A
Sequence 62530, A
Sequence 62532, A
Sequence 67774, A
Sequence 76811, A
Sequence 76977, A
Sequence 76978, A
Sequence 76, Appl
Sequence 839, App
Sequence 2046, Ap
Sequence 3096, Ap
Sequence 33928, A
```

```
; SOFTWARE: PatentIn version 3.3
```

```
; SEQ ID NO 102
; LENGTH: 182190
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-102
```

```
Query Match 87.2%; Score 21.8; DB 7; Length 182190;
Best Local Similarity 92.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAATCT 25
Db 49368 AAAAAAAAAATTCACAAAAATCT 49392
```

```
RESULT 2
US-11-121-086-9
; Sequence 9, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
```

```
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138, 6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9
; LENGTH: 196200
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-9
```

```
Query Match 87.2%; Score 21.8; DB 7; Length 196200;
Best Local Similarity 92.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAATCT 25
Db 67349 AAAAAAAAAATTCACAAAAATCT 67373
```

```
RESULT 3
US-11-121-086-10
; Sequence 10, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
```

```
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138, 6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10
; LENGTH: 199321
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-10
```

```
Query Match 87.2%; Score 21.8; DB 7; Length 199321;
Best Local Similarity 92.0%; Pred. No. 48;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAATCGCAACAATCT 25
Db 67349 AAAAAAAAAATTCACAAAAATCT 67373
```

Db 190751 AAAAAAAAAATCGCAAAACAAATCT 190775

RESULT 4

US-10-995-561-71043/c
; Sequence 71043, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71043
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-71043

Query Match 83.2%; Score 20.8; DB 6; Length 201;
Best Local Similarity 91.7%; Pred. No. 51;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAAAACAAATC 24

Db 70 AAAAAAAAAATTGAAACAAATC 47

RESULT 5

US-10-750-185-35318
; Sequence 35318, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MGI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35318
; LENGTH: 928
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-35318

Query Match 83.2%; Score 20.8; DB 6; Length 928;
Best Local Similarity 91.7%; Pred. No. 60;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAAAACAAATC 24

Db 170 AAAAAAAAAAGCAACAAATC 193

US-10-995-561-13421/c

; Sequence 13421, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF

; FILE REFERENCE: CL001559

; CURRENT APPLICATION NUMBER: US/10/995,561

; CURRENT FILING DATE: 2004-11-24

; NUMBER OF SEQ ID NOS: 85702

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13421

; LENGTH: 403278

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc_feature

; LOCATION: (1)...(403278)

; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-

US-10-995-561-13421

Query Match 83.2%; Score 20.8; DB 6; Length 403278;
Best Local Similarity 91.7%; Pred. No. 11e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAAAACAAATC 24

Db 73644 AAAAAAAAAATTGAAACAAATC 73621

RESULT 7

US-11-014-071-1
; Sequence 1, Application US/11014071
; Publication No. US20050246796A1
; GENERAL INFORMATION:
; APPLICANT: Cigan, Andrew M.
; APPLICANT: Fox, Timothy W.
; APPLICANT: Hershey, Howard P.
; APPLICANT: Unger, Erica
; APPLICANT: Wu, Yongzhong
; TITLE OF INVENTION: Dominant Gene Suppression Transgenes and
; FILE REFERENCE: 1554
; CURRENT APPLICATION NUMBER: US/11/014,071
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: 60/530,478
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: 60/591,975
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)...(1112)
; OTHER INFORMATION: P67
US-11-014-071-1

Query Match 81.6%; Score 20.4; DB 9; Length 1112;
Best Local Similarity 95.5%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAAAACAA 22

Db 752 AAAAAAAAAATCGCAACAAA 773

US-11-112-908-29

; Sequence 29, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole

```
APPLICANT: Davis, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
PRIOR FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 29
LENGTH: 131855
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-29
```

```
Query Match      81.6%; Score 20.4; DB 7; Length 131855;
Best Local Similarity 95.5%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAATCGCAACAA 22
DB      129641 AAAAAAAAAATCGCAACAA 129662
```

```
RESULT 9
US-11-112-908-30
Sequence 30, Application US/11112908
Publication No. US20050260659A1
GENERAL INFORMATION:
```

```
APPLICANT: Harris, Cole
APPLICANT: Davis, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 30
LENGTH: 143389
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-30
```

```
Query Match      81.6%; Score 20.4; DB 7; Length 143389;
Best Local Similarity 95.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAATCGCAACAA 22
DB      97414 AAAAAAAAAATCGCAACAA 97435
```

```
RESULT 10
US-10-857-780-3/C
Sequence 3, Application US/10857780
Publication No. US20050272043A1
GENERAL INFORMATION:
APPLICANT: ROTH, RICHARD B.
APPLICANT: BRAUN, ANDREAS
```

```
APPLICANT: KAMMERER, STEFAN M.
APPLICANT: NELSON, MATTHEW ROBERTS
APPLICANT: RYLAND, RICHARD HENRY
APPLICANT: ROYAL, WRIGHTSON, CAROLYN R.
TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
FILE REFERENCE: SEQ-4069-CP
CURRENT APPLICATION NUMBER: US/10/857,780
CURRENT FILING DATE: 2004-05-28
PRIOR APPLICATION NUMBER: 10/723,681
PRIOR FILING DATE: 2003-11-25
PRIOR APPLICATION NUMBER: 60/490,234
PRIOR FILING DATE: 2003-07-24
PRIOR APPLICATION NUMBER: 60/525,239
PRIOR FILING DATE: 2003-11-25
NUMBER OF SEQ ID NOS: 4962
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 147700
TYPE: DNA
ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (51510)..(51510)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (51526)..(51526)
OTHER INFORMATION: n is a, c, g, or t
US-10-857-780-3
```

```
Query Match      81.6%; Score 20.4; DB 6; Length 147700;
Best Local Similarity 95.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAATCGCAACAA 22
DB      75954 AAAAAAAAAATCGCAACAA 75933
```

```
RESULT 11
US-11-112-908-24
Sequence 24, Application US/11112908
Publication No. US20050260659A1
GENERAL INFORMATION:
```

```
APPLICANT: Harris, Cole
APPLICANT: Davis, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 24
LENGTH: 150314
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-24
```

```
Query Match      81.6%; Score 20.4; DB 7; Length 150314;
Best Local Similarity 95.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAAATCGCAACAA 22
```



```
Db      11881 AAAAAAAAAATCGCAAAAAA 11902
;
;
;
RESULT 12
; Sequence 28, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Hartis, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 28
; LENGTH: 166020
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-28

Query Match      81.6%; Score 20.4; DB 7; Length 166020;
Best Local Similarity 95.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCGCAACAA 22
;
;
;
Db      143895 AAAAAAAAAATCGCAAAAAA 143916
;
;
;
RESULT 13
US-10-995-561-28667
; Sequence 28667, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28667
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-28667

Query Match      80.8%; Score 20.2; DB 6; Length 201;
Best Local Similarity 88.0%; Pred. No. 82;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCGCAACAAATCT 25
;
;
;
Db      91 AAAAAAAAAATCGCAAAAAAATAT 115
;
;
;
RESULT 14
US-10-995-561-41795
; Sequence 41795, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42212
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-41797

Query Match      80.8%; Score 20.2; DB 6; Length 201;
Best Local Similarity 88.0%; Pred. No. 82;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCGCAACAAATCT 25
;
;
;
Db      48 AAAAAAAAAATCGCAAAAAAATCT 72
;
;
;
RESULT 15
US-10-995-561-41797
; Sequence 41797, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41797
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-41797

Query Match      80.8%; Score 20.2; DB 6; Length 201;
Best Local Similarity 88.0%; Pred. No. 82;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAATCGCAACAAATCT 25
;
;
;
Db      46 AAAAAAAAAATCGCAAAAAAATCT 70
;
;
;
RESULT 16
US-10-995-561-42212
; Sequence 42212, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42212
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-42212
```

```

; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

```

```

; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13286
; LENGTH: 1125000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1125000)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13286

Query Match          80.8%; Score 20.2; DB 6; Length 1125000;
Best Local Similarity 88.0%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 439944 AAAAAAAAAATCCCAAAATCT 439968

RESULT 22
US-10-750-185-41542
; Sequence 41542, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 41542
; LENGTH: 1423
; TYPE: DNA
; ORGANISM: Bovine 1986688092568
US-10-750-185-41542

Query Match          79.2%; Score 19.8; DB 6; Length 1423;
Best Local Similarity 91.3%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAAT 23
Db 1201 AAAAAAAAAAGCAAAAT 1223

RESULT 23
US-10-750-185-26307/c
; Sequence 26307, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```

```

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 26307
; LENGTH: 1674
; TYPE: DNA
; ORGANISM: Bovine 19866880351257
US-10-750-185-26307

Query Match          79.2%; Score 19.8; DB 6; Length 1674;
Best Local Similarity 91.3%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAAT 23
Db 30 AAAAAAAAAATCCCAACACAT 8

RESULT 24
US-10-750-185-50068/c
; Sequence 50068, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 50068
; LENGTH: 2126
; TYPE: DNA
; ORGANISM: Bovine 19866880634843
US-10-750-185-50068

Query Match          76.8%; Score 19.2; DB 6; Length 2126;
Best Local Similarity 87.5%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATC 24
Db 1589 AGAAAAAATTAATGCAACAAATC 1566

RESULT 25
US-10-750-185-57390/c
; Sequence 57390, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
```

```

; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57390
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Bovine 19866880506377
; US-10-750-185-57390

Query Match          76.8%; Score 19.2; DB 6; Length 2652;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATC 24
    |||||
Db 1071 AAAAAAAAAACCAACCAAC 1048

RESULT 26
; Sequence 53, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 176503
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-121-086-53

Query Match          76.8%; Score 19.2; DB 7; Length 176503;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATC 24
    |||||
Db 22162 AAAAAAAAAATGCAACCAAAAC 22185

RESULT 27
; US-11-121-086-53/C
; Sequence 53, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 176503
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-121-086-53

Query Match          76.8%; Score 19.2; DB 7; Length 176503;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATC 24
    |||||
Db 22162 AAAAAAAAAATGCAACCAAAAC 22185
```

```

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATC 24
    |||||
Db 129922 AAAAAAAAAATCGCAACAATC 129899

RESULT 28
; US-10-995-561-62120
; Sequence 62120, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62120
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-995-561-62120

Query Match          75.2%; Score 18.8; DB 6; Length 201;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATA 22
    |||||
Db 50 AAAAAAAAAAAGCAACAAACA 71

RESULT 29
; US-10-750-185-34921
; Sequence 34921, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34921
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Bovine 19866880875512
; US-10-750-185-34921

Query Match          75.2%; Score 18.8; DB 6; Length 1284;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATA 22
    |||||
Db 706 AAAAAAAAAAAGCAACAAACA 727

RESULT 30
; US-10-750-185-42879
```

```
; Sequence 42879, Application US/10750185
; Publication No. US2005026063A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42879
; LENGTH: 1481
; TYPE: DNA
; ORGANISM: Bovine 1986681758473
US-10-750-185-42879

Query Match          75.2%; Score 18.8; DB 6; Length 1481;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAATCGCAACAAA 22
    |||||
Db 31 AAAAAAAAAAATCGCAACAAA 52

RESULT 31
US-11-102-240-127
; Sequence 127, Application US/11102240
; Publication No. US20050260647A1
; GENERAL INFORMATION:
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: ANTIBODIES TO POLYPEPTIDES ENCODED BY A NUCLEIC ACID UNDEREXPRESSION
; FILE REFERENCE: P3230R1C106C
; CURRENT APPLICATION NUMBER: US/11/102,240
; CURRENT FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: 10/063662
; PRIOR FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: 10/006867
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 60/170262
; PRIOR FILING DATE: 199-12-09
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 127
; LENGTH: 1505
; TYPE: DNA
; ORGANISM: Homo Sapien
US-11-102-240-127

Query Match          75.2%; Score 18.8; DB 7; Length 1505;
Best Local Similarity 90.9%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAATCGCAACAAA 22
    |||||
Db 1475 AAAAAAAAAAATCGCAACAAA 1496

RESULT 32
```

```
US-10-995-561-13491/C
; Sequence 13491, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13491
; LENGTH: 23983
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13491

Query Match          75.2%; Score 18.8; DB 6; Length 23983;
Best Local Similarity 90.9%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAATCGCAACAAA 22
    |||||
Db 22681 AAAAAAAAAAATCGCAACAAA 22660

RESULT 33
US-11-121-086-96/C
; Sequence 96, Application US/11121086
; Publication No. US2005026459A1
; GENERAL INFORMATION:
; APPLICANT: NIELSEN, TIM S.
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 96
; LENGTH: 139054
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-96

Query Match          75.2%; Score 18.8; DB 7; Length 139054;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAATCGCAACAAA 22
    |||||
Db 57122 AAAAAAAAAAATCGCAACAAA 57101

RESULT 34
US-11-121-086-18
; Sequence 18, Application US/11121086
; Publication No. US2005026459A1
; GENERAL INFORMATION:
; APPLICANT: NIELSEN, TIM S.
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
```

```

; SEQ ID NO 18
; LENGTH: 175023
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-18

Query Match          75.2% Score 18.8; DB 7; Length 175023;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAA 22
DB 90165 AAAAAAAAAATCGCAACAAA 90186

RESULT 35
US-11-121-086-106/c
; Sequence 106, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 106
; LENGTH: 179777
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-106

Query Match          75.2% Score 18.8; DB 7; Length 179777;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAA 22
DB 120485 AAAAAAAAAATCGCAACAAA 120464

RESULT 36
US-11-121-086-37
; Sequence 37, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 37
; LENGTH: 184000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-37

Query Match          75.2% Score 18.8; DB 7; Length 184000;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAAA 22
DB 13725 AAAAAAAAAATCGCAACAAA 13746
```

```

RESULT 37
US-11-112-908-33/c
; Sequence 33, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 33
; LENGTH: 217623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-33

Query Match          75.2% Score 18.8; DB 7; Length 217623;
Best Local Similarity 90.9%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAAAAAAAAATCGCAACAAATCT 25
DB 138439 AAAAAAAAAATCGCAACAAATCT 138418

RESULT 38
US-10-995-561-13421
; Sequence 13421, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13421
; LENGTH: 403278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(403278)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13421

Query Match          75.2% Score 18.8; DB 6; Length 403278;
Best Local Similarity 90.9%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAAAAAAAAATCGCAACAAATCT 25
DB 109063 AAAAAAAAAATCGCAACAAATCT 109084

RESULT 39
US-10-995-561-17363/c
```

```
; Sequence 17363, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17363
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-17363

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 201;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 85 AAAAAAAAAAGCACAACACTCT 61

RESULT 40
US-10-995-561-31318/c
; Sequence 31318, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31318
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-31318

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 201;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 100 AAAAAAAAAATCAACCAATCT 76

RESULT 41
US-10-995-561-60450/c
; Sequence 60450, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60450
; LENGTH: 201
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-995-561-60450

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 201;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 39 AAAAAAAAAATCAAGCAATGCT 15

RESULT 42
US-10-995-561-60534/c
; Sequence 60534, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60534
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-60534

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 201;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 98 AAAAAAAAAATCAAGCAATGCT 74

RESULT 43
US-10-995-561-83119/c
; Sequence 83119, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83119
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-83119

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 201;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 47 AAAAAAAAAAGAAACAACTCT 23

RESULT 44
US-10-750-185-722
; Sequence 722, Application US/10750185
```

```

; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 722
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT06419
US-10-750-185-722

Query Match          74.4%; Score 18.6; DB 6; Length 600;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
DB 329 AAAAAAAAAAGCTGAAACAATCT 353

RESULT 45
US-10-750-185-3939
; Sequence 3939, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3939
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT08354
US-10-750-185-3939

Query Match          74.4%; Score 18.6; DB 6; Length 600;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
DB 19 AAAAAAAAAATTTCAACAATCT 43

RESULT 46
US-10-986-501-36
; Sequence 36, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.

```

```

; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: PZ013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 36
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-986-501-36

Query Match          74.4%; Score 18.6; DB 6; Length 606;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAACAATCT 25
DB 546 AAAAAAAAAATTTCAACAATTT 570

RESULT 47
US-10-750-185-55963/c
; Sequence 55963, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 55963
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Bovine 19866881392778
US-10-750-185-55963

Query Match          74.4%; Score 18.6; DB 6; Length 998;

```


Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 899 AAAAAAAAAATCTCCCAAAATCT 875

RESULT 48

US-10-750-185-53443
; Sequence 53443, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53443
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Bovine 19866880435748
US-10-750-185-53443

Query Match 74.4%; Score 18.6; DB 6; Length 1029;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 80 AAAAAAAAAACCCACACAAATCT 104

RESULT 49

US-10-750-185-36104/C
; Sequence 36104, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36104
; LENGTH: 1120
; TYPE: DNA
; ORGANISM: Bovine 19866881440381
US-10-750-185-36104

Query Match 74.4%; Score 18.6; DB 6; Length 1120;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 448 AAAAAAAAAATCTCAAGATATCT 424

RESULT 50

US-10-750-185-27139/C
; Sequence 27139, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27139
; LENGTH: 1205
; TYPE: DNA
; ORGANISM: Bovine 19866881791028
US-10-750-185-27139

Query Match 74.4%; Score 18.6; DB 6; Length 1205;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAACAAATCT 25
Db 897 AAAAAAAAAACACAAAAAATAT 873

Search completed: December 14, 2005, 11:40:02
Job time : 188.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-2

Perfect score: 25
Sequence: 1 aaaaaaaaaatcgacgacaaatc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ph:*
8: gb_ph:*
9: gb_ph:*
10: gb_ph:*
11: gb_ph:*
12: gb_ph:*
13: gb_ph:*
14: gb_ph:*
15: gb_ph:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.4	93.6	40899	8 CH19F14121	AD000090 Homo sapi
2	23.4	93.6	149964	8 AC104686	AC104686 Homo sapi
3	22.4	89.6	110000	14 AC152182_0	AC152182 Mus muscu
4	22.4	89.6	110000	14 AC152182_1	Continuation (2 of
5	22.4	89.6	110000	14 AC155165_1	Continuation (2 of
6	22.4	89.6	198822	9 AC158235	AC158235 Mus muscu
7	22.4	89.6	258911	14 AC107548	AC107548 Rattus no
8	21.8	87.2	37805	2 AF125964	AF125964 Caenorhab
9	21.8	87.2	49613	8 ALJ359186	ALJ359186 Human DNA
10	21.8	87.2	59765	8 AC0051179	AC0051179 Homo sapi
11	21.8	87.2	66331	14 AC087578	AC087578 Homo sapi
12	21.8	87.2	90007	8 ALI39234	ALI39234 Human DNA
13	21.8	87.2	92982	8 ALJ589642	ALJ589642 Human DNA
14	21.8	87.2	108736	8 CENS01RGR	ALJ59179 Human chr
15	21.8	87.2	111002	8 BX842242	BX842242 Homo sapi
16	21.8	87.2	111862	14 ALI39143	ALI39143 Homo sapi
17	21.8	87.2	112862	15 AC025781	AC025781 Arabidops
18	21.8	87.2	114835	8 AP001628	AP001628 Homo sapi

19	21.8	87.2	129402	8 AC011510	AC011510 Homo sapi
20	21.8	87.2	135754	8 AC125612	AC125612 Homo sapi
21	21.8	87.2	140938	14 AC121628	AC121628 Rattus no
22	21.8	87.2	142728	8 HSDJ792G4	AL049636 Human DNA
23	21.8	87.2	143324	8 ALI37881	ALI37881 Human DNA
24	21.8	87.2	151875	8 ALI36128	ALI36128 Human DNA
25	21.8	87.2	154957	8 AC019270	AC019270 Homo sapi
26	21.8	87.2	155523	14 AC025367	AC025367 Homo sapi
27	21.8	87.2	161659	14 AC147311	AC147311 Pan trogl
28	21.8	87.2	161841	8 AC135279	AC135279 Homo sapi
29	21.8	87.2	163237	8 AC130885	AC130885 Homo sapi
30	21.8	87.2	163554	14 AC021786	AC021786 Homo sapi
31	21.8	87.2	163731	8 AC008349	AC008349 Homo sapi
32	21.8	87.2	166517	14 ALJ591663	ALJ591663 Homo sapi
33	21.8	87.2	168833	14 AC021706	AC021706 Homo sapi
34	21.8	87.2	169993	8 ALJ353616	ALJ353616 Human DNA
35	21.8	87.2	171732	8 AC022844	AC022844 Homo sapi
36	21.8	87.2	172779	8 AL451107	AL451107 Human DNA
37	21.8	87.2	173491	14 AC073894	AC073894 Homo sapi
38	21.8	87.2	175134	14 ALI61623	ALI61623 Homo sapi
39	21.8	87.2	189791	14 AC152354	AC152354 Pan trogl
40	21.8	87.2	196831	14 AC068681	AC068681 Homo sapi
41	21.8	87.2	203753	8 AC104343	AC104343 Homo sapi
42	21.8	87.2	206249	9 AC165290	AC165290 Mus muscu
43	21.8	87.2	224450	14 AC016311	AC016311 Homo sapi
44	21.8	87.2	224450	14 AC016311	AC016311 Homo sapi
45	21.8	87.2	233195	14 AC115438	AC115438 Rattus no
46	21.8	87.2	242704	8 BS000230	BS000230 Pan trogl
47	21.8	87.2	280569	14 AC128115	AC128115 Rattus no
48	21.8	87.2	299744	14 AC024147	AC024147 Homo sapi
49	21.8	87.2	306130	14 AC110996	AC110996 Homo sapi
50	21.8	87.2	325493	14 AC107171	AC107171 Rattus no
51	21.8	87.2	340000	8 AP001747	AP001747 Homo sapi
52	21.4	85.6	175409	8 AC098971	AC098971 Homo sapi
53	21.4	85.6	211922	5 BX649578	BX649578 Zebrafish
54	20.8	83.2	14026	8 AC092209	AC092209 Homo sapi
55	20.8	83.2	68419	14 AC016014	AC016014 Homo sapi
56	20.8	83.2	74832	8 AP000455	AP000455 Homo sapi
57	20.8	83.2	84791	8 AC113340	AC113340 Homo sapi
58	20.8	83.2	108239	14 AC151550	AC151550 Dasyatis n
59	20.8	83.2	110000	15 AP008214_106	Continuation (107
60	20.8	83.2	110000	15 AP008214_308	Continuation (109
61	20.8	83.2	110000	15 AP008212_067	Continuation (68 o
62	20.8	83.2	110000	15 AP008212_067	Continuation (69 o
63	20.8	83.2	125057	9 AC108402	AC108402 Mus muscu
64	20.8	83.2	132427	15 AP003218	AP003218 Oryza sat
65	20.8	83.2	134599	15 AP004158	AP004158 Oryza sat
66	20.8	83.2	134599	8 AC011385	AC011385 Homo sapi
67	20.8	83.2	142057	4 AC091297	AC091297 Sus scro
68	20.8	83.2	147263	14 AC142718	AC142718 Macaca mu
69	20.8	83.2	151668	15 AP003509	AP003509 Oryza sat
70	20.8	83.2	152010	14 CR936486	CR936486 Dario rer
71	20.8	83.2	161277	15 AP002972	AP002972 Oryza sat
72	20.8	83.2	162753	14 AC144197	AC144197 Macaca mu
73	20.8	83.2	163408	14 AC091389	AC091389 Homo sapi
74	20.8	83.2	165363	8 HS459L4	AL031120 Human DNA
75	20.8	83.2	169212	8 AP002392	AP002392 Homo sapi
76	20.8	83.2	178088	14 AC073645	AC073645 Homo sapi
77	20.8	83.2	201139	9 AC122865	AC122865 Mus muscu
78	20.8	83.2	201399	9 AC1357149	AC1357149 Human DNA
79	20.8	83.2	214619	9 AC122267	AC122267 Mus muscu
80	20.8	83.2	217817	14 AC145402	AC145402 Gorilla g
81	20.8	83.2	226546	14 AC157049	AC157049 Bos tauru
82	20.8	83.2	237547	14 AC144198	AC144198 Macaca mu
83	20.8	83.2	248489	14 AC162662	AC162662 Bos tauru
84	20.8	83.2	340000	8 HS21C018	AL163218 Homo sapi
85	20.8	83.2	349980	6 AX344574	AX344574 Sequence
86	20.4	81.6	7300	8 AB000462	AB000462 Homo sapi
87	20.4	81.6	22970	8 HS1247F6	Z68279 Human DNA
88	20.4	81.6	25485	5 CR391942	CR391942 Zebrafish
89	20.4	81.6	54965	14 AC014556	AC014556 Drosophila
90	20.4	81.6	149969	8 HS1323A24	AL121750 Human DNA
91	20.4	81.6	160111	14 AP000780	AP000780 Homo sapi

92	20.4	81.6	161614	2	AC011706	Drosophila
93	20.4	81.6	163969	8	AP001138	Homo sapi
94	20.4	81.6	172558	8	AC093580	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
95	20.4	81.6	174820	2	AC010917	Drosophila
96	20.4	81.6	185267	14	AC165999	Otolemur
97	20.4	81.6	195987	9	AC122537	Mus muscu
98	20.4	81.6	196476	14	AC055890	Drosophila
99	20.4	81.6	198074	14	CR376775	Danio rer
100	20.4	81.6	199255	14	AC084195	Homo sapi
101	20.4	81.6	200482	14	AC164923	Otolemur
102	20.4	81.6	203904	14	AC157040	Bos tauru
103	20.4	81.6	204847	8	BS000066	Pan trogl
104	20.4	81.6	208122	14	AC094413	Rattus no
105	20.4	81.6	215677	14	AC127061	Rattus no
106	20.4	81.6	223082	14	AC127061	Rattus no
107	20.4	81.6	226601	9	AC122863	Mus muscu
108	20.4	81.6	266430	5	BX324132	Zebrafish
109	20.4	81.6	266753	5	CR381619	Zebrafish
110	20.4	81.6	300051	2	AE003570	Drosophila
111	20.4	81.6	340000	8	AP001680	Homo sapi
112	20.2	80.8	382	8	AB012172	Homo sapi
113	20.2	80.8	459	10	BV241629	S234P6211
114	20.2	80.8	460	10	BV304791	S236P6314
115	20.2	80.8	547	10	BV191793	sgm17208
116	20.2	80.8	601	10	BV186114	sgm14996
117	20.2	80.8	609	10	BV373272	S231P6598
118	20.2	80.8	615	10	BV621513	S215P6603
119	20.2	80.8	704	2	AY753166	Plasmodiu
120	20.2	80.8	704	2	AY753167	Plasmodiu
121	20.2	80.8	704	2	AY753168	Plasmodiu
122	20.2	80.8	735	10	BV478495	S591P5712
123	20.2	80.8	760	10	BV629695	S216P6062
124	20.2	80.8	784	2	AY575008	Plasmodiu
125	20.2	80.8	862	10	BV467903	G591P6000
126	20.2	80.8	1193	2	PVY18842	Plasmodium
127	20.2	80.8	1390	15	BT002114	Arabidops
128	20.2	80.8	1534	15	AY084253	Arabidops
129	20.2	80.8	1546	15	AY136395	Arabidops
130	20.2	80.8	1825	9	PRRLPM62X	Deer mouse
131	20.2	80.8	2535	1	AF239256	Acinetoba
132	20.2	80.8	3875	6	C0612968	Sequence
133	20.2	80.8	6162	6	AX356490	Sequence
134	20.2	80.8	7145	6	AX279989	Sequence
135	20.2	80.8	7145	6	AX356409	Sequence
136	20.2	80.8	7145	6	AX251751	Sequence
137	20.2	80.8	7928	6	AX344165	Sequence
138	20.2	80.8	7928	6	AX344997	Sequence
139	20.2	80.8	7928	6	AX348556	Sequence
140	20.2	80.8	9311	8	AC148675	Macaca mu
141	20.2	80.8	10565	2	AF442041	Plasmodiu
142	20.2	80.8	11186	13	AF632332	Coccoloba
143	20.2	80.8	25140	8	AC020824	Homo sapi
144	20.2	80.8	29105	6	AX695335	Sequence
145	20.2	80.8	29118	8	HS310H5	Human DNA
146	20.2	80.8	32167	8	AY549314	Homo sapi
147	20.2	80.8	34956	8	AY398667	Homo sapi
148	20.2	80.8	35642	8	AL603825	Human DNA
149	20.2	80.8	39188	8	AC004754	Homo sapi
150	20.2	80.8	41936	6	AX335752	Sequence

ALIGNMENTS

RESULT 1
 CH19P14121/c 40899 bp linear PRI 22-MAR-1997
 LOCUS Homo sapiens DNA from chromosome 19q13.1 cosmid f14121 containing
 DEFINITION ATP4A and GADPH-2 genes, genomic sequence.
 ACCESSION
 VERSION Add000090.1 GI:1905894
 KEYWORDS H,K-ATPase (ATP4A); chromosome 19; glyceraldehyde 3-phosphate dehydrogenase.

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 40899)
TITLE	Lamerdin, J.E.
JOURNAL	Submitted (07-NOV-1996) J.E. Lamerdin, Human Genome Center, Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA, USA, 94551 janeacgt.llnl.gov owlornak.llnl.gov
COMMENT	Human Genome Center Biology and Biotechnology Research Program Lawrence Livermore National Laboratory 7000 East Avenue Livermore, CA 94550 USA.
FEATURES	Location/Qualifiers
source	1..40899
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="19"
	/clone="f14121"
	/cell_line="UV5HL9-5B"
	/cell_type="fibroblast"
	/clone_id="L19NC02 F2 chromosome 19-specific cosmid library"
repeat_region	6..234
	/note="repeat match = HSL02759; putative"
	/rpt_family="Alu"
repeat_region	161..253
	/note="repeat match = HSL04058; putative"
	/rpt_family="Alu"
exon	1088..1178
	/note="predicted exon: similarity to M. musculus testis-specific isoform of glyceralde hyde 3- phosphate dehydrogenase (frame 2); putative"
	/number=1
misc_feature	1088..1178
	/note="predicted exon, grail2exons_human_1.3; frame=2, forward strand, quality=good; putative"
misc_feature	1089..1178
	/note="similarity: gl 567204; gl 567204 (M60978) glyceraldehyde 3-phosphate dehydrogenase [Mus musculus] 1259..1377
exon	1259..1377
	/note="predicted exon: similarity to M. musculus testis-specific isoform of glyceralde hyde 3- phosphate dehydrogenase (frame 2); putative"
	/number=2
misc_feature	1259..1377
	/note="predicted exon, grail2exons_human_1.3; frame=1, forward strand, quality=excellent; putative"
misc_feature	1262..1375
	/note="similarity: gl 567204; gl 567204 (M60978) glyceraldehyde 3-phosphate dehydrogenase [Mus musculus] 1687..1795
misc_feature	1687..1795
	/note="predicted exon, grail2exons_human_1.3; frame=1, forward strand, quality=good; putative"
misc_feature	1712..1801
	/note="similarity: gl 567204; gl 567204 (M60978) glyceraldehyde 3-phosphate dehydrogenase [Mus musculus] 1714..1795
exon	1714..1795
	/note="predicted exon: similarity to M. musculus testis-specific isoform of glyceralde hyde 3- phosphate dehydrogenase (frame 2); putative"
	/number=3
misc_feature	2109..2287
	/note="predicted exon, grail2exons_human_1.3; frame=2, forward strand, quality=excellent; putative"

```

misc_feature      2109..2261
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
. . . putative"
exon              2109..2260
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=4
exon              2434..2596
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=5
misc_feature      2434..2596
/note="predicted exon, grail2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"
2435..2599
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
. . . putative"
misc_feature      3678..3775
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
. . . putative"
exon              3678..3775
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=6
misc_feature      3678..3775
/note="predicted exon, grail2exons human_1.3; frame=2,
forward strand, quality=excellent; putative"
3867..3939
/note="predicted exon: similarity to M. musculus
testis-specific isoform of glyceralde hyde 3- phosphate
dehydrogenase (frame 2); putative"
/number=7
misc_feature      3867..3939
/note="predicted exon, grail2exons human_1.3; frame=0,
forward strand, quality=excellent; putative"
3868..3936
/note="similarity: gi|567204; gi|567204 (M60978)
glyceraldehyde 3-phosphate dehydrogenase [Mus musculus]
. . . putative"
exon              4509..4585
/note="predicted exon (exon 1): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=1
misc_feature      4509..4585
/note="predicted exon, grail2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"
4657..4726
/note="predicted exon (exon 2): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=2
misc_feature      4657..4726
/note="predicted exon, grail2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"
5295..5354
/note="predicted exon (exon 3): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=3
misc_feature      5295..5354
/note="predicted exon, grail2exons human_1.3; frame=2,
forward strand, quality=good; putative"
5441..5577
/note="predicted exon (exon 4): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=4
misc_feature      5441..5577
/note="predicted exon, grail2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"

```

```

exon              5714..5798
/note="predicted exon (exon 5): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=5
exon              5915..6009
/note="predicted exon (exon 6): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=6
misc_feature      5921..6009
/note="predicted exon, grail2exons human_1.3; frame=1,
forward strand, quality=excellent; putative"
6093..6216
/note="predicted exon (exon 7): similarity to C. elegans
ZK418.5 gene product (U00047); putative"
/number=7
misc_feature      6093..6216
/note="predicted exon, grail2exons human_1.3; frame=0,
forward strand, quality=good; putative"
complement(9384..22397)
/number=1
CDS
gene
complement(9384..22397)
/gene="ATP4a"
complement(9384..22397)
10216..10349,11806..11951,12475..12598,13691..13845,
13925..14103,14209..14359,14445..14581,15681..15856,
16423..16615,17130..17264,17345..17454,17761..17959,
18573..18841,19131..19383,19587..19700,21203..21406,
21977..22036,22152..22295,22386..22397))
/function="human gastric H,K-ATPase catalytic subunit"
/note="human gastric H,K-ATPase catalytic subunit gene
(ATP4a), inverse strand; putative"
/codon_start=1
/product="human gastric H,K-ATPase catalytic subunit"
/protein_id="AAB50172.1"
/db_xref="GI:1905895"
/translation="MGKAENYELYSVELSPGPGDMAAMSKKKKGGGGRKKELE
MKKEKINHOISVALEKQKYSATKGSASIAAVALLDGDKNALRPPTGYVYK
PARLAGLGLMMVAAALICIAIRIQASBDLTDDNLTALIALIAVVVITGCGYV
QEFKSTINIASFKLVPOQATVIRDGDFOINADQLVGDVLEKGGRRVADIRILA
AOGKGVNSLITGSESPOTRSPCTHSPLTRIAPFSTICLGTVGLVAVNGDRT
IIGKIASIASGVENEKTPIDALEIFPDVILINGLILPFAATPFIVAMCFLTMAMV
FMAIVAVVPEGLATVTCISLAKSLASNGCVKNALEVTGSTVICSIDKGTIL
TONMTVSHLMPDNHIDTDEDSOGTPOSESTWALRCVLTLCRAAFKSGODA
VPVKRIVIGDASTALKKSELTLSNMGCRDPRPKCEIPNSTKFOUSITLED
PRDRHLILVMKAPERVLDERSSITLIGQELPLDEQNBARQTVLISLGGERYLGF
CQYLINRKYDPPGAFVDEANNFPSSGCEFGALVSMIDPPRATVPDAVLKCRGTGIRV
IMVYGDHPTAKAIAASVGLISBGSSEGTVAARVAVDVONRKRDAVACVINGWOLK
MDPSEIVEALRTHPEMVFPARTSPQKLVIVESCORLGAIVAVTGDVNDSPALKKAD

```

Query Match 93.6%; Score 23.4; DB 8; Length 40899;
 Best Local Similarity 96.0%; Pred. No. 96;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACACAAATCT 25
 Db 25504 AAAAAAAAAATCGAACAAATCT 25480

RESULT 2
 AC104686 149964 bp DNA linear PRI 13-MAY-2005
 LOCUS AC104686 BAC clone Rpl1-74M11 from 4, complete sequence.
 DEFINITION Homo sapiens BAC clone Rpl1-74M11 from 4, complete sequence.
 ACCESSION AC104686 AC024968
 VERSION AC104686.2 GI:18855136
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 149964)
 AUTHORS Boyer E., Haakenson, W. and Radionenko, M.
 TITLE The sequence of Homo sapiens BAC clone Rpl1-74M11
 JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 149964)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 149964)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA (bases 1 to 149964)
 REFERENCE 4 (bases 1 to 149964)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAY-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Feb 21, 2002 this sequence version replaced gi:17933860.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0074M11
 Drafting Center: WIBR

NOTICE:

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:

The RRC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,B., Tachio,M., Catalane,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.reagen.com) or Pieter de Jong and coworkers at http://www.chori.org

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC005699, 2000 bp overlap; the clone sequenced to the right is RP11-168E17. Actual end of this clone is at base position 149964 of RP11-74M11.

Data from AC024335 was used to finish this clone, AC024968.

The sequence of AC024968 has been incorporated into AC104686.

FEATURES

source
 location/Qualifiers
 1. 149964
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="RP11-74M11"
 /clone_lib="RRC1-11"

ORIGIN

Query Match 93.6%, Score 23.4, DB 8, Length 149964;

Best Local Similarity 96.0%; Pred. No. 71;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 AAAAAAAAAATCGACAAATCT 25
 Db 123240 AAAAAAAAAATCGACAAATCT 123216

RESULT 3

AC152182_0

WPCOMMENT

Sequence split into 5 fragments LOCUS AC152182 Accession AC152182

Fragment Name	Begin	End
AC152182_0	1	110000
AC152182_1	100001	210000
AC152182_2	200001	310000
AC152182_3	300001	410000
AC152182_4	400001	452855

LOCUS AC152182 452855 bp DNA linear HTG 07-DEC-2004
 DEFINITION Mus musculus chromosome 16 clone RP23-12286, WORKING DRAFT
 SEQUENCE, 26 unordered pieces.

ACCESSION AC152182 GI:56404175
 VERSION AC152182.3
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 452855)
 AUTHORS Wilson,R.K.
 TITLE The sequence of Mus musculus clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 452855)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (29-OCT-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 452855)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (07-DEC-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Dec 7, 2004 this sequence replaced gi:56236317.

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: M_BA0122E06

----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 441530 bases at least Q40
 Consensus quality: 443831 bases at least Q30
 Consensus quality: 445190 bases at least Q20

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 26 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1669: contig of 1669 bp in length
 * 1670 1769: gap of unknown length

```
* 1770 3329: contig of 1560 bp in length
* 3330 3429: gap of unknown length
* 3430 5881: contig of 2452 bp in length
* 5882 5981: gap of unknown length
* 5982 8941: contig of 2960 bp in length
* 8942 9041: gap of unknown length
* 9042 12275: contig of 3224 bp in length
* 12276 12375: gap of unknown length
* 12376 16214: contig of 3839 bp in length
* 16215 16314: gap of unknown length
* 16315 23272: contig of 6958 bp in length
* 23273 23372: gap of unknown length
* 23373 30507: contig of 7135 bp in length
* 30508 30607: gap of unknown length
* 30608 40567: contig of 9960 bp in length
* 40568 40667: gap of unknown length
* 40668 49453: contig of 8786 bp in length
* 49454 49553: gap of unknown length
* 49554 58920: contig of 9367 bp in length
* 58921 59020: gap of unknown length
* 59021 68861: contig of 9841 bp in length
* 68862 68961: gap of unknown length
* 68962 78889: contig of 9928 bp in length
* 78890 78989: gap of unknown length
* 78990 92676: contig of 13687 bp in length
* 92677 92776: gap of unknown length
* 92777 104713: contig of 11937 bp in length
* 104714 104813: gap of unknown length
* 104814 116360: contig of 11547 bp in length
* 116361 116460: gap of unknown length
* 116461 132663: contig of 16203 bp in length
* 132664 132763: gap of unknown length
* 132764 146020: contig of 13257 bp in length
* 146021 146120: gap of unknown length
* 146121 159570: contig of 13450 bp in length
* 159571 159670: gap of unknown length
* 159671 180442: contig of 20772 bp in length
* 180443 180542: gap of unknown length
* 180543 204769: contig of 24227 bp in length
* 204770 204869: gap of unknown length
* 204870 232406: contig of 27537 bp in length
* 232407 232506: gap of unknown length
* 232507 264122: contig of 31616 bp in length
* 264123 264222: gap of unknown length
* 264223 316904: contig of 52682 bp in length
* 316905 317004: gap of unknown length
* 317005 364926: contig of 47922 bp in length
* 364927 365026: gap of unknown length
* 365027 452855: contig of 87829 bp in length.
```

FEATURES

```
source
1..452855
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="16"
  /clone="RP23-122E6"
misc_feature
1..1669
  /note="assembly_name:Contig13"
gap
1670..1769
  /estimated_length=unknown
misc_feature
1770..3329
  /note="assembly_name:Contig14"
gap
3330..3429
  /estimated_length=unknown
misc_feature
3430..5881
  /note="assembly_name:Contig16"
gap
5882..5981
  /estimated_length=unknown
misc_feature
5982..8941
  /note="assembly_name:Contig17"
gap
8942..9041
  /estimated_length=unknown
misc_feature
9042..12275
```

```
/note="assembly_name:Contig18"
12276..12375
  /estimated_length=unknown
misc_feature
12376..16214
  /note="assembly_name:Contig19"
16215..16314
  /estimated_length=unknown
misc_feature
16315..23272
  /note="assembly_name:Contig20"
23273..23372
  /estimated_length=unknown
misc_feature
23373..30507
  /note="assembly_name:Contig21"
30508..30607
  /estimated_length=unknown
misc_feature
30608..40567
  /note="assembly_name:Contig22"
40568..40667
  /estimated_length=unknown
misc_feature
40668..49453
  /note="assembly_name:Contig23"
49454..49553
  /estimated_length=unknown
misc_feature
49554..58920
  /note="assembly_name:Contig24"
58921..59020
  /estimated_length=unknown
misc_feature
59021..68861
  /note="assembly_name:Contig25"
68862..68961
  /estimated_length=unknown
misc_feature
68962..78889
  /note="assembly_name:Contig26"
78890..78989
  /estimated_length=unknown
misc_feature
78990..92676
  /note="assembly_name:Contig27"
92677..92776
  /estimated_length=unknown
misc_feature
92777..104713
  /note="assembly_name:Contig28"
104714..104813
  /estimated_length=unknown
misc_feature
104814..116360
  /note="assembly_name:Contig29"
116361..116460
  /estimated_length=unknown
misc_feature
116461..132663
  /note="assembly_name:Contig30"
132664..132763
  /estimated_length=unknown
misc_feature
132764..146020
  /note="assembly_name:Contig31"
146021..146120
  /estimated_length=unknown
misc_feature
146121..159570
  /note="assembly_name:Contig32"
159571..159670
  /estimated_length=unknown
misc_feature
159671..180442
  /note="assembly_name:Contig33"
180443..180542
  /estimated_length=unknown
misc_feature
180543..204769
```

```
Query Match      89.6%; Score 22.4; DB 14; length 110000;
Best Local Similarity 95.8%; Pred: No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 AAAAAAAAAATCGACGACAAATCT 25
DB 105229 AAAAAAAAAATCAGCAAAATCT 105252
```

RESULT 4
AC152182_1
WPCOMMENT

Sequence split into 5 fragments LOCUS AC152182 Accession AC152182

Fragment Name	Begin	End
AC152182_0	1	110000
AC152182_1	100001	210000
AC152182_2	200001	310000
AC152182_3	300001	410000
AC152182_4	400001	452855

Continuation (2 of 5) of AC152182 from base 100001 (AC152182 Mus musculus chromosome 16

Query Match 89.6%; Score 22.4; DB 14; Length 110000;
Best Local Similarity 95.8%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGACGACAAATCT 25
Db 5229 AAAAAAAAAATCGACGACAAATCT 5252

RESULT 5
AC155165_1
WPCOMMENT

Sequence split into 6 fragments LOCUS AC155165 Accession AC155165

Fragment Name	Begin	End
AC155165_0	1	110000
AC155165_1	100001	210000
AC155165_2	200001	310000
AC155165_3	300001	410000
AC155165_4	400001	510000
AC155165_5	500001	551847

Continuation (2 of 6) of AC155165 from base 100001 (AC155165 Mus musculus chromosome 16

Query Match 89.6%; Score 22.4; DB 14; Length 110000;
Best Local Similarity 95.8%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGACGACAAATCT 25
Db 12459 AAAAAAAAAATCGACGACAAATCT 12482

RESULT 6
AC158235/c
LOCUS AC158235 198822 bp DNA linear ROD 30-JUL-2005
DEFINITION Mus musculus chromosome 16 clone RP23-185P17, complete sequence.
ACCESSION AC158235
VERSION AC158235.3 GI:71533487
KEYWORDS HTG:
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 198822)
Wilson, R.K.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 198822)
Wilson, R.K.
Direct Submission
Submitted (09-MAR-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 198822)
Wilson, R.K.
Direct Submission
Submitted (13-MAR-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 198822)
Wilson, R.K.
Direct Submission

JOURNAL
COMMENT

Submitted (30-JUL-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 30, 2005 this sequence version replaced gi:161656452.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
Project Information
Center project name: M.BA0185P17

FEATURES
source

location/Qualifiers
1..198822
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="16"
/clone="RP23-185P17"

ORIGIN

Query Match 89.6%; Score 22.4; DB 9; Length 198822;
Best Local Similarity 95.8%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAATCGACGACAAATCT 25
Db 121120 AAAAAAAAAATCGACGACAAATCT 121097

RESULT 7
AC107548
LOCUS

AC107548 258911 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-44B14, *** SEQUENCING IN PROGRESS
***, 14 unordered pieces.

ACCESSION AC107548
VERSION AC107548.5 GI:30580718
KEYWORDS HTG; HTGS_PHASER; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS

Wuzny, D., Martie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederick, D.,
Deigado, O., Denison, S., Deramo, C., Ding, Y., Din, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Becotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Gante, R., Garcia, A., Garner, T., Garza, M.,
Georgescu, E., Geier, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunsarune, P., Healand, W., Hamli, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hu, Y., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louie, H., Lozano, R., Lutz, X., Ma, J.,
Maheshwari, M., Mahindartine, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidaa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newcom, N., Nguyen, N., Norris, S., Nwokilemeh, O., Okwou, G., Olarnunagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Pindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quito, J., Rachin, B., Reeves, K., Regier, M.A., Reigh, R., Rellis, B., Rellis, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Unpublished
2 (bases 1 to 258911)
Worley, K.C.

Direct Submission
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 258911)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23268120.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GLCXX
Center clone name: CH230-44B14

----- Summary Statistics
Assembly program: Atlas 3.0
Consensus quality: 245471 bases at least Q40
Consensus quality: 248216 bases at least Q30
Consensus quality: 250042 bases at least Q20
Estimated insert size: 254905; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

FEATURES
source
* be preserved.
1 232716: contig of 232716 bp in length
* 232717 232816: gap of unknown length
* 232817 234030: contig of 1214 bp in length
* 234031 234130: gap of unknown length
* 234131 235586: contig of 1456 bp in length
* 235587 235686: gap of unknown length
* 235687 237582: contig of 1896 bp in length
* 237583 237682: gap of unknown length
* 237683 238979: contig of 1297 bp in length
* 238980 239079: gap of unknown length
* 239080 240419: contig of 1340 bp in length
* 240420 240519: gap of unknown length
* 240520 242180: contig of 1661 bp in length
* 242181 242280: gap of unknown length
* 242281 244290: contig of 2010 bp in length
* 244291 244390: gap of unknown length
* 244391 246035: contig of 1645 bp in length
* 246036 246135: gap of unknown length
* 246136 248426: contig of 2291 bp in length
* 248427 248526: gap of unknown length
* 248527 249594: contig of 1068 bp in length
* 249595 249694: gap of unknown length
* 249695 253025: contig of 3331 bp in length
* 253026 253125: gap of unknown length
* 253126 256274: contig of 3149 bp in length
* 256275 256375 258911: gap of unknown length
* 256375 258911: contig of 2537 bp in length.

Location/Qualifiers
1. 258911
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-44B14"
1. 1502
/note="wgs end extension
clone_end:5p6"
4677. 5504
/note="clone boundary
clone_end:5p6
site:EcORI
end sequence: BH291676"
complement(230271..231097)
/note="clone boundary
clone_end:T7
site:EcORI
end sequence: BH291675"
232717..232816
/estimated_length=unknown
234031..234130
/estimated_length=unknown
235587..235686
/estimated_length=unknown
237583..237682
/estimated_length=unknown
238980..239079
/estimated_length=unknown
240420..240519
/estimated_length=unknown
242181..242280
/estimated_length=unknown
244291..244390
/estimated_length=unknown
246036..246135
/estimated_length=unknown
248427..248526
/estimated_length=unknown
249595..249694
/estimated_length=unknown
253026..253125
/estimated_length=unknown
256275..256374
/estimated_length=unknown

ORIGIN

Query Match 89.6%; Score 22.4; DB 14; Length 256911;
 Best Local Similarity 95.8%; Pred. No. 1.4e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCCGACAAATC 24
 |||||
 DB 69804 AAAAAAAAAATCCGACAAATC 69827

RESULT 8 AF125964 37805 bp DNA linear INV 22-SEP-2004
 LOCUS Caenorhabditis elegans cosmid W03G1, complete sequence.
 ACCESSION AF125964
 VERSION AF125964.1 GI:4262631
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita;
 Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 37805)
 WormBase Consortium
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 9851916
 2 (bases 1 to 37805)
 Pauley, A., Scheet, P. and Harper, M.
 The sequence of C. elegans cosmid W03G1
 unpublished (2001)
 3 (bases 1 to 37805)
 Waterston, R.
 Direct Submission
 Submitted (04-FEB-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 4 (bases 1 to 37805)
 Waterston, R.
 Direct Submission
 Submitted (24-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 5 (bases 1 to 37805)
 Waterston, R.
 Direct Submission
 Submitted (28-AUG-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 37805)
 Waterston, R.
 Direct Submission
 Submitted (21-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 37805)
 Waterston, R.
 Direct Submission
 Submitted (19-APR-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 8 (bases 1 to 37805)
 Wilson, R.
 Direct Submission
 Submitted (15-JUN-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 9 (bases 1 to 37805)
 Wilson, R.
 Direct Submission
 Submitted (15-MAY-2004) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

REFERENCE
 AUTHORS WormBase Consortium
 CONSRMT Direct Submission
 TITLE Submitted (22-SEP-2004) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 JOURNAL Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: submissions@wustl.edu and jess@sanger.ac.uk

COMMENT
 NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.
 This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m3 subclone.
 For a graphical representation of this clone sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=W03G1;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is Y65D7A, 200 bp overlap; the 3' clone is F09C11, 200
 bp overlap. Actual start of this clone is at base position 1 of
 W03G1, actual end is at 37609 of W03G1.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 GeneFINDER (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yujin Kohara
 (http://www.ddj.nig.ac.jp/c-elegans/hm1/CE_INDEX.html) and The C.
 elegans ORFeome cloning project (<http://wordfidd.dcel.harvard.edu/>),
 similarity to other proteins from BlastX analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans GenBank submissions,
 and personal communications with C. elegans researchers. tRNAs
 are predicted using the program tRNAscan-SB (Lowe, T.M. and
 Eddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).
 Location/Qualifiers

1..37805
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="W03G1"
 complement(2183..3598)
 /gene="W03G1.5"
 /locus_tag="W03G1.5"
 complement(2183..3598)
 /gene="W03G1.5"
 /locus_tag="W03G1.5"
 /standard_name="W03G1.5"
 /note="contains similarity to Homo sapiens Hornerin,
 ENSMBL:ENSP0000032953, coded for by the following C.
 elegans cDNAs: CKS76560, CKS77835, CKS78135,
 CKS78138, CKS78332, CKS78474, CKS81667, CKS81825,
 CKS81923, CKS88488, OSTR139G5_1, OSTR139G5_2"

FEATURES
 source
 gene
 CDS

Query Match	Similarity	Score	DB	Length	37805;
Best Local	92.0%	Pred. No. 3.7e+02;			
Matches	23;	Conservative	0;	Mismatches	2;
				Indels	0;
				Gaps	0;

1 AAAAAAAAAATCGCAGACAAATCT 25

[illegible]

gene

short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
join(41239..41453,42428..42509)
/gene="EIF2C4"

mRNA

/locus_tag="RP11-239B6.1-002"
join(41239..41453,42428..42509)
/gene="EIF2C4"

gene

/product="eukaryotic translation initiation factor 2C, 4"
/note="match: ESTs: Em:AW295680.1"
join(complement(ALJ54864.16:5815)..6078),4633..4798,
10613..10733,13064..13245,13430..13566,13677..13811,
14499..14586,19178..19325,19563..19682,19808..19936,
20188..20321,21741..21925,23585..23744,28916..29231,
29367..29501,37908..38009,38605..38804,41248..45641)
/gene="EIF2C4"

mRNA

/locus_tag="RP11-239B6.1-001"
join(complement(ALJ54864.16:5815)..6078),4633..4798,
10613..10733,13064..13245,13430..13566,13677..13811,
14499..14586,19178..19325,19563..19682,19808..19936,
20188..20321,21741..21925,23585..23744,28916..29231,
29367..29501,37908..38009,38605..38804,41248..45641)
/gene="EIF2C4"

CDS

/product="eukaryotic translation initiation factor 2C, 4"
/note="match: ESTs: Em:AA482095.1 Em:AA651785.1
Em:A1022110.1 Em:A1674879.1 Em:BF979523.1 Em:BQ214750.1
Em:BQ77272.1 Em:BX103141.1
match: CDNAs: Em:AB046787.1 Em:AB081474.1 Em:AK00040.1
Em:AK030018.1 Em:AK032475.1 Em:AL31908.1"
join(complement(ALJ54864.16:5815)..5833),4633..4798,
10613..10733,13064..13245,13430..13566,13677..13811,
14499..14586,19178..19325,19563..19682,19808..19936,
20188..20321,21741..21925,23585..23744,28916..29231,
29367..29501,37908..38009,38605..38804,41248..41356)
/gene="EIF2C4"

/locus_tag="RP11-239B6.1-002"
/standard_name="OTTHMP000000431"
/note="match: proteins: Sw:O9U118 Tr:O8BTF4 Tr:O8CJF8
Tr:O9HCK5"

/codon_start=1
/product="eukaryotic translation initiation factor 2C, 4"
/protein_id="CAH71584.1"
/db_xref="GI:55664190"
/db_xref="InterPro:IPR003100"
/db_xref="InterPro:IPR003165"

/translation="WEALGPGPPASLFLQPPRPGLGVGKPIRLANHFQVLPKIDY
YHYVDI.KPEKRPVRVNEVDVTWRHFKQI.FDROPGYGRKMYTAHPLPIGRD
VDMEVTLKPEBKDOTFKVSVQVSVSLQLLALAGHLNEVPPDSVOALDVTIRHLP
SMRYTPGSRSPFEGYHPLGGREYVGFHOSVR.PAMMMMLNIDVATATFRRD
PLIFMCEVLDIININQTKELTDSQRYKPTKEIRGLKNEVTTCQMKRKYRVCNTR
RPASHQTFPQLDENGQAMECTVAQTFKQKTSLOLKYRPLPCLOVQGEQKTYLRELC
NIVAGORCIKTLNDQSTMTIKATARSAPROEBSISLVKSNVAGGPDPLKKEGIV
VHNMTBLTGRVLNAPMLQYGRNKTVATPQGVQWDRKQSVYGEIKIWAVACFPAP
QKOCREDLKSFTDLRKISKDAMPIDGQPCFCYQAGASVPMFKHLMKTVGLO
LIVVILPGKTPVAEVRKVDPTLGMATQCYQVNVKTSQPTLSN.CIKTNATLGGI
NNV.VPHORPSVPOQVPIFLGADVTHPAGGKPSIAAVGSDGHSRSCATVRQ
TSRBOISOBELYSQVETODLTNMYRELILQYKSTRFKRPIILYIRGSGVSKQKQVA
WPELAIKAKCISLEEDYRPIGITIVQKRHTLPCADTKERKGVNAGPAGTVTS
TITRPSDFPLCSHAGIQGSRPSHVQVMDNDGCFPADELTQLTLTLCYVCTGS
VSI.PAPAYALVAFRARIYHLVDKDHDSAEQSHVSGQNGRDPALAKAVQIHHDTQ
TMFA"

polyA_signal
polyA_site
misc_feature

45620..45625
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-001"
45641
/gene="EIF2C4"
/locus_tag="RP11-239B6.1-001"
47614
/note="Clone_left_end: RP4-789D17"

ORIGIN

Query Match

87.2%; Score 21.8; DB 8; Length 49613;

Best Local Similarity 92.0%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGAGCAATCT 25
Db 34925 AAAAAAAAAATCGAGCAATCT 34949

RESULT 10
AC005179/c
LOCUS
DEFINITION
Homo sapiens chromosome 5, p1 clone 1041F10 (LBNL H88), complete
sequence.
ACCESSION
AC005179 AC001127 AC001128 AC001129 AC001130 AC001131 AC001578
AC001579 AC002267
VERSION
AC005179.1 GI:3258607
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 59765)
Kimmerly, M., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Kadner, K., Miguel, T., Miller, T., Plutnick, S., Pollard, M.,
Rojesti, H., Subramanian, S. and Martin, C.H.
Sequencing of human chromosome 5
2 (bases 1 to 59765)
Ricke, D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
3 (bases 1 to 59765)
Kimmerly, M., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Plutnick, S., Pollard, M.,
Rojesti, H., Subramanian, S. and Martin, C.H.
Direct Submission
Submitted (26-JUN-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source
Location/Qualifiers
1..59765
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q"
/clone="P1 1041F10"
/note="LBNL H88"
56..354
/rpt_family="Alu"
complement(join(882..1069,25942..25994))
/note="Genscan prediction 1"
1312..1635
/rpt_family="Alu"
complement(join(1987..2055,22602..22663))
/note="81 & 87% identity AL022345"
complement(2422..2661)
/rpt_family="Alu"
complement(2991..3286)
/rpt_family="Alu"
complement(3865..4174)
/rpt_family="Alu"
3872..3893
/note="(T)22"
/rpt_type=tandem
/rpt_unit="c"
4178..4215
/note="(GT)19"
/rpt_type=tandem

```

repeat_region      /rpt untc="gt"      complement(4240. .4540)
                    /rpt_family="MER42"
repeat_region      complement(4739. .4803)
                    /rpt_family="MER42"
repeat_region      complement(4865. .5216)
                    /rpt_family="Alu"
repeat_region      complement(5288. .5407)
                    /rpt_family="MER20"
prim_transcript    join(5861. .6093,6630. .7108)
                    /standard_name="rs10b04"
prim_transcript    /note="99$ identity EST dbEST:AA251303 & AA251631"
                    complement(5861. .6139)
                    /standard_name="ztc94g06"
repeat_region      /note="100$ identity EST dbEST:AA416991 & AA470056"
                    7364. .7693
                    /rpt_family="Alu"
repeat_region      7764. .8072
                    /rpt_family="Alu"
repeat_region      complement(8156. .8469)
                    /rpt_family="Alu"
repeat_region      8496. .8861
                    /rpt_family="Alu"
                    complement(8658. .9027)
repeat_region      /rpt_family="Alu"
                    9051. .9258
                    /note="98$ identity EST yq02c04.61"
misc_feature        /db_xref=dbEST:R89075"
                    9450. .9843
misc_feature        /note="96$ identity EST56708"
                    /db_xref=dbEST:T33112"
                    9742. .10037
repeat_region      /rpt_family="Alu"
                    complement(10540. .11010)
prim_transcript    /standard_name="zv80e10"
                    /note="94$ & 98$ identity dbEST:AA424065"
                    10798. .10843
repeat_region      /note="AC) 23"
                    /rpt_type=tandem
                    /rpt_untc="ac"
                    complement(10872. .11053)
repeat_region      /rpt_family="MER46"
                    complement(11904. .12023)
repeat_region      /rpt_family="L1"
                    12279. .12775
repeat_region      /rpt_family="MLT2B2"
                    complement(13756. .14122)
repeat_region      /rpt_family="MER4"
                    complement(14564. .14767)
repeat_region      /rpt_family="Alu"
                    14761. .14822
repeat_region      /rpt_family="L1"
                    complement(14827. .15152)
repeat_region      /rpt_family="Alu"
                    15326. .15463
repeat_region      /rpt_family="L1"
                    16193. .16372
repeat_region      /rpt_family="MST"
                    16462. .16580
repeat_region      /rpt_family="MST"
                    complement(17170. .17478)
repeat_region      /rpt_family="Alu"
                    18021. .18365
repeat_region      /rpt_family="Alu"
                    complement(18437. .18736)
repeat_region      /rpt_family="Alu"
                    19075. .19192
repeat_region      /rpt_family="MER42"
                    19234. .19539
repeat_region      /rpt_family="Alu"
                    complement(19793. .20087)
repeat_region      /rpt_family="Alu"
                    20588. .20718

```

```

misc_feature      /rpt_family="Alu"
                  complement(21091. .21368)
                  /note="98% identity EST z139h02.s1"
                  /db_xref="dbEST:AA676208"
repeat_region     22529. .22866
                  /rpt_family="Alu"
                  22832. .22852
                  /note=" (A) 21"
                  /rpt_type=tandem
                  /rpt_unit="a"
repeat_region     complement(23446. .23659)
                  /rpt_family="MER58A"
                  24216. .24546
                  /rpt_family="Alu"
                  24505. .24524
                  /note=" (A) 20"
                  /rpt_type=tandem
                  /rpt_unit="a"
repeat_region     24548. .24787
                  /rpt_family="L1"
                  25966. .26285
                  /rpt_family="Alu"
                  complement(26552. .26847)
                  /rpt_family="Alu"
                  complement(27205. .27504)
                  /rpt_family="Alu"
                  complement(27522. .27746)
                  /rpt_family="MER30"
                  28653. .28806
                  /rpt_family="MIR"
                  29240. .29607
                  /note="98% identity B36828 BAC end sequence"
repeat_region     30284. .30374
                  /rpt_family="LINE2"
                  32813. .33177
                  /standard_name="endogenous retrovirus"
                  33453. .33574
                  /rpt_family="Alu"
                  33661. .33829
                  /rpt_family="Alu"
                  34077. .34716
                  /rpt_family="Alu"
                  35152. .35345
                  /rpt_family="MLT1"
                  complement(35525. .35660)
                  /rpt_family="Alu"
                  complement(35726. .35889)
                  /rpt_family="Alu"
                  35958. .36290
                  /rpt_family="Alu"
                  complement(37374. .37494)
                  /note="GAAT 2 excellent exon, frame 2"
repeat_region     complement(38108. .38761)
                  /rpt_family="Alu"
                  38772. .39090
                  /rpt_family="Alu"
                  39065. .39085
                  /note=" (A) 21"
                  /rpt_type=tandem
                  /rpt_unit="a"
repeat_region     complement(39706. .39996)
                  /rpt_family="Alu"
                  40686. .41001
                  /rpt_family="Alu"

```

Query Match	87.2%	Score 21.8	DB 8	Length 59765
Similarity	92.0%	Pred. No. 3.3e+02		
Local	Conservative	0	Mismatches 2	Indels 0
1	AAAAAAAAATCGACGCAATCT	25		
DB	27642 AAAAAAAAAATCGCAAAAAAAAAATCT	27618		

QY	1	AAAAAAAAAATCGCAGACAATCT	25
Db	27642	AAAAAAAAAAATCGCAAAAAAATCT	27618

RESULT 11
AC087578/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-116P9 map 11, LOW-PASS
SEQUENCE SAMPLING.
AC087578
AC087578.2 GI:1184093
VERSION
KEYWORDS HTG; HTGS_PHASE0
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-116P9
JOURNAL Unpublished
REFERENCE
AUTHORS 2 (bases 1 to 66331)
TITLE
JOURNAL
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Baerlein, V., Boguslavskiy, L., Bouckgalter, B., Brown, A.,
Cammarata, T., Campianno, A., Choepel, T., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karata, A., LaRocque, K., Lamasares, R., Landers, T.,
Lehoczy, T., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPeckers, R., Meldrum, J., Menus, L., Mhova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Sances, R., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Sougnen, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strausen, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (10-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2001 this sequence version replaced gi:12061453.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I12293
Center clone name: 116_P_9

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 689: contig of 689 bp in length
* 690 789: gap of 100 bp
* 790 1485: contig of 626 bp in length
* 1486 1585: gap of 100 bp
* 1586 2277: contig of 692 bp in length
* 2278 2377: gap of 100 bp

2378 3019: contig of 642 bp in length
* 3020 3119: gap of 100 bp
* 3120 3811: contig of 692 bp in length
* 3812 3911: gap of 100 bp
* 3912 4598: contig of 687 bp in length
* 4599 4698: gap of 100 bp
* 4699 5377: contig of 679 bp in length
* 5378 5477: gap of 100 bp
* 5478 6166: contig of 689 bp in length
* 6167 6266: gap of 100 bp
* 6267 6972: contig of 706 bp in length
* 6973 7072: gap of 100 bp
* 7073 7769: contig of 697 bp in length
* 7770 7869: gap of 100 bp
* 7870 8570: contig of 701 bp in length
* 8571 8670: gap of 100 bp
* 8671 9376: contig of 706 bp in length
* 9377 9476: gap of 100 bp
* 9477 10156: contig of 680 bp in length
* 10157 10256: gap of 100 bp
* 10257 10956: contig of 700 bp in length
* 10957 11056: gap of 100 bp
* 11057 11757: contig of 701 bp in length
* 11758 11857: gap of 100 bp
* 11858 12546: contig of 689 bp in length
* 12547 12646: gap of 100 bp
* 12647 13359: contig of 713 bp in length
* 13360 13459: gap of 100 bp
* 13460 14155: contig of 696 bp in length
* 14156 14255: gap of 100 bp
* 14256 14947: contig of 692 bp in length
* 14948 15047: gap of 100 bp
* 15048 15725: contig of 678 bp in length
* 15726 15825: gap of 100 bp
* 15826 16543: contig of 718 bp in length
* 16544 16643: gap of 100 bp
* 16644 17346: contig of 703 bp in length
* 17347 17446: gap of 100 bp
* 17447 18148: contig of 702 bp in length
* 18149 18248: gap of 100 bp
* 18249 18984: contig of 736 bp in length
* 18985 19084: gap of 100 bp
* 19085 19792: contig of 708 bp in length
* 19793 19892: gap of 100 bp
* 19893 20618: contig of 726 bp in length
* 20619 20718: gap of 100 bp
* 20719 21403: contig of 685 bp in length
* 21404 21503: gap of 100 bp
* 21504 22201: contig of 698 bp in length
* 22202 22301: gap of 100 bp
* 22302 23001: contig of 700 bp in length
* 23002 23101: gap of 100 bp
* 23102 23812: contig of 711 bp in length
* 23813 23912: gap of 100 bp
* 23913 24617: contig of 705 bp in length
* 24618 24717: gap of 100 bp
* 24718 25419: contig of 702 bp in length
* 25420 25519: gap of 100 bp
* 25520 26266: contig of 747 bp in length
* 26267 26366: gap of 100 bp
* 26367 27065: contig of 699 bp in length
* 27066 27165: gap of 100 bp
* 27166 27880: contig of 715 bp in length
* 27881 27980: gap of 100 bp
* 27981 28678: contig of 698 bp in length
* 28679 28778: gap of 100 bp
* 28779 29500: contig of 722 bp in length
* 29501 29600: gap of 100 bp
* 29601 30292: contig of 692 bp in length
* 30293 30392: gap of 100 bp
* 30393 31116: contig of 724 bp in length
* 31117 31216: gap of 100 bp
* 31217 31940: contig of 724 bp in length

```

* 31941 32040: gap of 100 bp
* 32041 32743: contig of 703 bp in length
* 32843 32843: gap of 100 bp
* 32844 33545: contig of 702 bp in length
* 33546 33645: gap of 100 bp
* 33646 34332: contig of 687 bp in length
* 34333 34432: gap of 100 bp
* 34433 35141: contig of 709 bp in length
* 35142 35241: gap of 100 bp
* 35242 35942: contig of 701 bp in length
* 35943 36042: gap of 100 bp
* 36043 36736: contig of 694 bp in length
* 36737 36836: gap of 100 bp
* 36837 37536: contig of 700 bp in length
* 37537 37637: gap of 100 bp
* 37637 38326: contig of 690 bp in length
* 38327 38426: gap of 100 bp
* 38427 39134: contig of 708 bp in length
* 39135 39234: gap of 100 bp
* 39235 39921: contig of 687 bp in length
* 39922 40021: gap of 100 bp
* 40022 40728: contig of 707 bp in length
* 40729 40828: gap of 100 bp
* 40829 41527: contig of 699 bp in length
* 41528 42319: contig of 692 bp in length
* 42320 42419: gap of 100 bp
* 42420 43111: contig of 692 bp in length
* 43112 43211: gap of 100 bp
* 43212 43911: contig of 700 bp in length
* 43912 44011: gap of 100 bp
* 44012 44706: contig of 695 bp in length
* 44707 44806: gap of 100 bp
* 44807 45524: contig of 718 bp in length
* 45525 45624: gap of 100 bp
* 45625 46319: contig of 695 bp in length
* 46320 46419: gap of 100 bp
* 46420 47118: contig of 699 bp in length
* 47119 47218: gap of 100 bp
* 47219 47925: contig of 707 bp in length
* 47926 48025: gap of 100 bp
* 48026 48721: contig of 696 bp in length
* 48722 48821: gap of 100 bp
* 48822 49515: contig of 694 bp in length
* 49516 49615: gap of 100 bp
* 49616 50311: contig of 696 bp in length
* 50312 50411: gap of 100 bp
* 50412 51113: contig of 702 bp in length
* 51114 51213: gap of 100 bp
* 51214 51917: contig of 704 bp in length
* 51918 52017: gap of 100 bp
* 52018 52720: contig of 703 bp in length
* 52721 52820: gap of 100 bp
* 52821 53489: contig of 669 bp in length
* 53490 53589: gap of 100 bp
* 53590 54278: contig of 689 bp in length
* 54279 54378: gap of 100 bp

```

```

Query Match 87.2% Score 21.8; DB 14; Length 66311;
Best Local Similarity 92.0%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAATCGACAAATCT 25
Db 8049 AAAAAAAAAATCGACAAATCT 8035

```

```

RESULT 12
AL139234 AL139234 90007 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP3-438D16 on chromosome Xq24-26.1
Contains the PDCD8 gene for programmed cell death 8
(apoptosis-inducing factor) (AIF), the RAB33A gene for RAB33A.

```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT

member RAS oncogene family (S10, Rabs10), the 3' end of a novel gene (FLJ20095) and two Cpg Islands, complete sequence.
AL139234
AL139234.19 GI:14787430
HTG: AIF; Cpg island; FLJ20095; PDCD8; RAB33A; Rabs10; S10.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 90007)
Bird.C.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Jul 17, 2001 this sequence version replaced gi:13751266.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP3-438D16 is from the library RP3-3 constructed by the group of P. Pierce de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

FEATURES

source

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

Location/Qualifiers

1..90007
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="q24-26.1"
/clone="RP3-438D16"
/clone_lib="RP3-3"
/complement(join(7540..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,16771..16879,17973..18049,18711..18795,23657..23747,25671..25801,25930..26054,27647..27746,33334..33464,43728..43929))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-002"
/complement(join(7540..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,16771..16879,17973..18049,18711..18795,23657..23747,25671..25801,25930..26054,27647..27746,33334..33464,43728..43929))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-002"
/product="programmed cell death 8 (apoptosis-inducing factor)"
/note="match: ESTs: BM927047.1 BY250260.1 BY263326.1 BY281562.1 H15605.1
match: CDNA: AF262320.1 AL049703.1"
complement(join(7540..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,16771..16879,17973..18049,18711..18795,23657..23747,25671..25801,25930..26054,27647..27746,33334..33464,43728..43929))

25671..25801,25930..26054,27647..27746,34638..34780,43728..43900))
/locus_tag="RP3-438D16.2-001"
/gene="PDCD8"
complement(join(7540..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,16771..16879,17973..18049,18711..18795,23657..23747,25671..25801,25930..26054,27647..27746,34638..34780,43728..43900))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-001"
/product="programmed cell death 8 (apoptosis-inducing factor)"
/note="match: ESTs: AL543045.1 AL544137.1 AL545607.1 AL550236.1 AL550264.1 AL557331.1 AL571252.1 AM518630.1 BE887067.1 BE903161.1 BE906258.1 BF027150.1 BF038249.1 BF182905.1 BG108121.1 BG115219.1 BG115901.1 BG116918.1 BG119647.1 BG437720.1 BG744018.1 BG744689.1 B1762702.1 B1826051.1 B1835393.1 BM013834.1 BM015333.1 BM471177.1 BM548814.1 BM807180.1 BM809802.1 BM921817.1 BQ058888.1 BQ062329.1 BQ065495.1 BQ066492.1 BQ212050.1 BQ300813.1 BQ308753.1 BQ581468.1 BQ161623.1 BQ165860.1 BQ170548.1 BQ178593.1 BQ517321.1 BQ552706.1 BX31768.1 BX341156.1 match: CDNAs: AF100928.1 AF375656.1 AF529274.1 AK000775.1 complement(join(7540..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,43728..43857))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
complement(join(7540..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,43728..43857))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
/product="programmed cell death 8 (apoptosis-inducing factor)"
/note="match: CDNAs: AL049704.1"
complement(join(7540..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,43728..43857))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
complement(join(7735..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,43728..43833))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
/standard_name="OTTHUMP0000024009"
/codon_start=1
/product="programmed cell death 8 (apoptosis-inducing factor)"
/protein_id="CA142779.1"
/db_xref="GI:57208843"
/translation="MFRCGGLAAGALKOKLVLTVCVRSPPQRNRLPARAGTEVOLFPEKGMGKILPEVLSNMTMEKVRREGVKVMNAIVQSVGSGLILKQGRKVEITDTHVAAGLEPVEIATGGLTETDSDPGCPRVNAELQASNTIVADDAACFPDIKGRRVREHDAVVGSRILAGENNTGAARPYHOSMNSMGLDGVGTAEALGLVDSILPTVGVFAVATADNPKSATGSGTGRSESESEASEITTPSTPAVPAQPVQGEYGVKGVIFVFLRDVVVGIWMNIFNRMPIARKIINDGQHEIDNEVAKLFNIED"
complement(join(7735..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,16771..16879,17973..18049,18711..18795,23657..23747,25671..25801,25930..26054,27647..27746,34638..34780,43728..43833))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
/standard_name="OTTHUMP0000024008"
/codon_start=1
/product="programmed cell death 8 (apoptosis-inducing factor)"
/protein_id="CA142780.1"
/db_xref="GI:57208844"

/translation="MFRCGGLAAGALKOKLVLTVCVRSPPQRNRLPARAGTEVOLFPEKGMGKILPEVLSNMTMEKVRREGVKVMNAIVQSVGSGLILKQGRKVEITDTHVAAGLEPVEIATGGLTETDSDPGCPRVNAELQASNTIVADDAACFPDIKGRRVREHDAVVGSRILAGENNTGAARPYHOSMNSMGLDGVGTAEALGLVDSILPTVGVFAVATADNPKSATGSGTGRSESESEASEITTPSTPAVPAQPVQGEYGVKGVIFVFLRDVVVGIWMNIFNRMPIARKIINDGQHEIDNEVAKLFNIED"
complement(join(7735..7806,8148..8344,9853..9977,11491..11633,14223..14363,14821..14909,15256..15363,16771..16879,17973..18049,18711..18795,23657..23747,25671..25801,25930..26054,27647..27746,34638..34780,43728..43833))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-003"
/standard_name="OTTHUMP0000024007"
/note="match: protein: O95831"
/codon_start=1
/product="programmed cell death 8 (apoptosis-inducing factor)"
/protein_id="CA142778.1"
/db_xref="GI:57208842"
/db_xref="InterPro:IPR001100"
/db_xref="InterPro:IPR001327"
/translation="MFRCGGLAAGALKOKLVLTVCVRSPPQRNRLPGNLFQRMHVPLELOMTRMASSGASGKIDNSVTVLVIGLSGAAYAAKTKEDKRNERSISGLPTEBOCKKALASGESEVPQKASHVFLIGGTAFAAARISIRAPDARVLIIVSDPELPVVRPPLSKELWFSDDPNTKTIRFQONGKERSITYFPPSPFVSADLP HIENGVAVLTGKVOLDVDRDNNVKINDGQITVEKCLTATGTPRSLSAIDBAQPE VKERTTTPRKIGDFRSLEKISRYSKSTITTIQGFSGSEALARKKARALATETVQLF PEKNMGKILPEVLSNMTMEKVRREGVKVMNAIVQSVGSGLILKQGRKVEITD HIIVAAVGLPEPVEIATGGLTETDSDPGCPRVNAELQASNTIVADDAACFPDIKGR RVREHDAVVGSRILAGENNTGAARPYHOSMNSMGLDGVGTAEALGLVDSILPTVGV FAKATADNPKSATGSGTGRSESESEASEITTPSTPAVPAQPVQGEYGVKGVIF YLRDQVYVGIWMNIFNRMPIARKIINDGQHEIDNEVAKLFNIED"
complement(join(14876..14909,15256..15896))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-004"
complement(join(14876..14909,15256..15896))
/gene="PDCD8"
/locus_tag="RP3-438D16.2-004"
/product="programmed cell death 8 (apoptosis-inducing factor)"
/note="match: ESTs: AI834231.1 BG675614.1 BG985116.1"
49826..63047
/gene="RAB33A"
/locus_tag="RP3-438D16.1-001"
Query Match 87.2%; Score 21.8; DB 8; Length 90007;
Best Local Similarity 92.0%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Cy 1 AAAAAAAAAATCCAGCAAACTCT 25
Db 36899 AAAAAAAAAATCCAGCAAAATCT 36923
RESULT 13
LOCUS AL589642
DEFINITION Human DNA sequence from clone Rp11-5554 on chromosome 9 Contains a TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa (TAF1) (CCG1, P250) pseudogene, two novel genes and a heterogeneous nuclear ribonucleoprotein C (C1/C2) (HNRPC) pseudogene, complete sequence.
ACCESSION AL589642 AC011902
VERSION AL589642.6 GI:16973047
KEYWORDS HTG; C1/C2; CCG1; HNRPC; P250; TAF1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 92982)

Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:16116510.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL, Sw: SWISSPROT, Tr: TrEMBL, Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-555J4 is from the library RPCT-11.2 constructed by the group of Pieper de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
Draft Sequence Produced by Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
http://genome.wustl.edu/gsc/index.shtml.

FEATURES
source
1..92982
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-555J4"
/clone_1ib="RPCT-11.2"
2000
/note="Clone_right_end: RP11-205M20"
complement(14164..19948)
/locus_tag="RP11-555J4.1-001"
/pseudo
complement(14164..19948)
/locus_tag="RP11-555J4.1-001"
/note="match: proteins: O35361 O97068 P21675 P51123 Q60544 Q9WTX0"
/psedo
/codon_start=1
join(17823..19424,32075..33054)
/locus_tag="RP11-555J4.4-001"
join(17823..19424,32075..33054)
/locus_tag="RP11-555J4.4-001"
/note="match: cDNAs: AK091653.1"
complement(join(27472..28112,29198..29395))
/locus_tag="RP11-555J4.3-001"
complement(join(27472..28112,29198..29395))
/locus_tag="RP11-555J4.3-001"
/note="match: ESTs: BU161407.1"
41508..41551
/note="Sequence from clone PCR only."
complement(60437..60889)
/locus_tag="RP11-555J4.2-001"

CDS
complement(60437..60889)
/locus_tag="RP11-555J4.2-001"
/note="match: proteins: AAH08243 AAM28214 P09234 P90815 Q03369 Q62241 Q875K5 Q9VE17 Q9ZB19"

misc_feature
/pseudo
/codon_start=1
78694..78748
/note="Sequence from clone PCR only."
90983
/note="Clone_left_end: RP11-462B18"

misc_feature
90983

ORIGIN
Query Match 87.2%; Score 21.8; DB 8; Length 92982;
Best Local Similarity 92.0%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACCAATCT 25
|||||
Db 52816 AAAAAAAAAATCGACCAATCT 52840
|||||

RESULT 14
CNS01RGR/c
LOCUS
DEFINITION
Human chromosome 14 DNA sequence BAC C-2282E14 of library CalTech-D
from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION
AL159179
KEYWORDS
HTG
AL159179.3 GI:11611160
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 108736)
Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P., Brothier,P., Catolico,L., Barbe,V., Pelletier,E., Artiguenave,F., Levy,M., Eckenberg,R., Bruls,T., deBardinas,V., Cruaud,C., Gysay,G., Saurin,W. and Weissenbach,J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 108736)
Genoscope.
Direct Submission
Submitted (26-APR-2001) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
On Dec 9, 2000 this sequence version replaced gi:7340700.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: segref@genoscope.cns.fr

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-862P13
Downstream BAC (overlapping the SP6 end) : R-125H8 (AC=AL135978)
----- Summary Statistics
Assembly program: Phrap; Version 2.0
Quality coverage: 10.0% in Q20 bases; sum-of-contigs

Overall quality chart :
Range : bases
0 : 1
1 - 9 : 9
10 - 19 : 40
20 - 29 : 215
30 - 39 : 2438
40 - 49 : 5878
50 - 59 : 4679
60 - 69 : 4679

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p34.1-34.3"
/clone="RP4-555P23"
/clone_lib="RPCT-4"
1. .6110
/note="assembly fragment:00533
fragment chain:1"
6211. .12365
/note="assembly fragment:00488
fragment chain:1"
12466. .14908
/note="assembly fragment:00991
fragment chain:1"
15009. .17724
/note="assembly fragment:00166
fragment chain:2"
17825. .25035
/note="assembly fragment:00257
fragment chain:2"
25136. .27936
/note="assembly fragment:00608
fragment chain:3"
28037. .31938
/note="assembly fragment:00316
fragment chain:3"
32039. .44844
/note="assembly fragment:00611
fragment chain:4"
44945. .57247
/note="assembly fragment:00630
fragment chain:4"
57348. .63813
/note="assembly fragment:00206.0"
63914. .70362
/note="assembly fragment:00382"
70463. .80246
/note="assembly fragment:00429"
80347. .94000
/note="assembly fragment:00431"
94101. .101709
/note="assembly fragment:00467"
101810. .107002
/note="assembly fragment:00663"
107103. .111862
/note="assembly fragment:00860"

ORIGIN
Query Match 87.2%; Score 21.8; DB 14; Length 111862;
Best Local Similarity 92.0%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGACACAATCT 25
Db 77558 AAAAAAAAAATCACAACAATCT 77582
RESULT 17
LOCUS AC025781 112862 bp DNA linear PLN 19-JUN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC F15C21 genomic sequence,
complete sequence.
ACCESSION AC025781
VERSION AC025781.8 GI:12322475
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (baaes 1 to 112862)

AUTHORS
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.U.,
Wu,D., Maiti,R., Roming,C.M., Koo,H., Fujii,C.Y., Utechtack,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Frazer,C.M.
Arabidopsis thaliana chromosome 1 BAC F15C21 genomic sequence
Unpublished
2 (baaes 1 to 112862)
Lin,X. and Kaul,S.
Direct Submission
Submitted (14-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (baaes 1 to 112862)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JUN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:11280766.
Address all correspondence to:atetigr.org

COMMENT
BAC clone F15C21 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mt.edu/GENSCAN.html), GenemarkEM (Mark Borodovsky,
http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmer.htm/glimmerm.html, and
Gensplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by repeatmasker (Arjan Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES

source

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F15C21"
/ecotype="Columbia"
1. .7871
/gene="F15C21.1"
/note="similar to GI:1100253 from [Arabidopsis thaliana]"
join(<1..24,120..307,390..440,515..554,643..767,844..960,
1036..1110,1187..1243,1361..1465,1577..1663,1761..1952,
2027..2215,2295..2405,2505..2594,2677..2853,2944..3021,
3097..3219,3293..3562,3666..3836,3858..3997,4077..4314,
4395..4747,4828..5041,5137..5313,5401..7551,7654..7871)
/gene="F15C21.1"
join(<1..24,120..307,390..440,515..554,643..767,844..960,
1036..1110,1187..1243,1361..1465,1577..1663,1761..1952,
2027..2215,2295..2405,2505..2594,2677..2853,2944..3021,
3097..3219,3293..3562,3666..3836,3858..3997,4077..4314,
4395..4747,4828..5041,5137..5313,5401..7551,7654..7710)
/gene="F15C21.1"
/codon_start=3
/product="acetyl-CoA carboxylase, putative, 5' partial;
1-7710"
/protein_id="AAGS1250.1"
/db_xref="GI:12322477"

gene

mRNA

CDS

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F15C21"
/ecotype="Columbia"
1. .7871
/gene="F15C21.1"
/note="similar to GI:1100253 from [Arabidopsis thaliana]"
join(<1..24,120..307,390..440,515..554,643..767,844..960,
1036..1110,1187..1243,1361..1465,1577..1663,1761..1952,
2027..2215,2295..2405,2505..2594,2677..2853,2944..3021,
3097..3219,3293..3562,3666..3836,3858..3997,4077..4314,
4395..4747,4828..5041,5137..5313,5401..7551,7654..7710)
/gene="F15C21.1"
/codon_start=3
/product="acetyl-CoA carboxylase, putative, 5' partial;
1-7710"
/protein_id="AAGS1250.1"
/db_xref="GI:12322477"
/translation="IPLMQIPERIRRFYIEHGQYDSWRKTSVVAFPDFPKAQSIRP
KGVCAVAVTSEDDDFKPSGVQVSEIFSKSPVAVAFVSKGCGIHSPDSQFGH
VFARESALAIANNVGLKQIRGEIRTVNDYITDLHNSDVRDNKIRHGMDSRI
AMRVRAERPMYLSVGGALYKASATSAVVDVVGLEKQIRPKHISLVHISOVSLN

REFERENCE 4 (bases 1 to 135754)
 AUTHORS Worley K.C.
 TITLE Direct Submission
 JOURNAL Submitted (30-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 135754)
 AUTHORS Worley K.C.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Sep 30, 2002 this sequence version replaced gi:22901969.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

SOURCE 1.135754
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /clone="RP11-117N2"
 /complement(1..3336)
 /note="Overlap bases 1..3336 of clone AC021052"
 /function="clone overlap"
 39..328
 repeat_region /rpt_family="AluSx"
 complement(820..1021)
 repeat_region /rpt_family="MIR"
 complement(1030..1329)
 repeat_region /rpt_family="AluJc"
 1540..1563
 repeat_region /rpt_family="TTTAA)n"
 2183..2480
 repeat_region /rpt_family="AluSc"
 complement(2723..3091)
 repeat_region /rpt_family="L1ME1"
 complement(3144..3452)
 repeat_region

/rpt_family="AluSx"
 3937..4053
 repeat_region /rpt_family="MLT1K"
 4181..4271
 repeat_region /rpt_family="MLT1K"
 complement(4272..4570)
 repeat_region /rpt_family="AluSx"
 4571..4645
 repeat_region /rpt_family="MLT1K"
 complement(4663..4819)
 repeat_region /rpt_family="MIR"
 5283..5585
 repeat_region /rpt_family="AluSp"
 5595..5904
 repeat_region /rpt_family="AluSp"
 5915..6218
 repeat_region /rpt_family="AluSx"
 6382..6650
 repeat_region /rpt_family="AluSx"
 6654..6690
 repeat_region /rpt_family="(CAAAA)n"
 complement(7736..7915)
 repeat_region /rpt_family="MER5A"
 8106..8268
 repeat_region /rpt_family="MER103"
 complement(8438..8736)
 repeat_region /rpt_family="AluSx"
 8779..9032
 repeat_region /rpt_family="AluSx"
 9035..9063
 repeat_region /rpt_family="(TAAA)n"
 9064..9182
 repeat_region /rpt_family="FLAM_C"
 9186..9211
 repeat_region /rpt_family="(TAAA)n"
 9277..9581
 repeat_region /rpt_family="AluSg"
 complement(9626..9753)
 repeat_region /rpt_family="AluSg/x"
 11338..11381
 repeat_region /rpt_family="CCCCCG)n"
 11669..11703
 repeat_region /rpt_family="GC_rich"
 14667..14704
 repeat_region /rpt_family="AT_rich"
 15189..15216
 repeat_region /rpt_family="(TTTG)n"
 complement(15985..16362)
 repeat_region /rpt_family="MLT1A2"
 complement(16405..16566)
 repeat_region /rpt_family="MIR"
 16779..16803
 repeat_region /rpt_family="AT_rich"
 16804..17094
 repeat_region /rpt_family="AluSx"
 17095..17147
 repeat_region /rpt_family="AT_rich"
 complement(17155..17441)
 repeat_region /rpt_family="AluSx"

Query Match 87.2% Score 21.8; DB 8; Length 135754;

Best Local Similarity 92.0%; Pred. No. 2.7e+02; Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACGACAAATCT 25
 |||||
 Db 121951 AAAAAAAAAATCGACAAAAATCT 121927

RESULT 21
 AC121628 140938 bp DNA linear HTG 15-NOV-2002
 LOCUS AC121628
 DEFINITION Rattus norvegicus clone CH230-106D2, WORKING DRAFT SEQUENCE.

[illegible]


```

/standard_name="OTTHUMP00000017135"
/codon_start=1
/protein_id="CAI23002.1"
/db_xref="GI:56204340"
/db_xref="InterPro:IPR007327"
/db_xref="UniProt/TREMBL:Q9BUC6"
/translation="MEAOAGLLETEPLOGCTDEDAVASADPSSMLSEEEKELAEV
QLEDEITLROVLSAKERHVLVEIKQKGLGMNLMNELKONFSKSMHDMOTTAYKKTHET
LSHGQKATAAFSNVGTAKSKFGDMSTISHSISMPMRRK"
join(127425..127443,AL121938.10:41776..41891,
AL121938.10:50816..50964,AL121938.10:69980..70081,
AL121938.10:84532..84541)
/gene="TPD52L1"
/locus_tag="RPI-16705.1-004"
/standard_name="OTTHUMP00000017136"
/codon_start=1
/protein_id="CAI23003.1"
/db_xref="GI:56204341"
/db_xref="InterPro:IPR007327"
/db_xref="UniProt/TREMBL:Q9C054"
/translation="MEAOAGLLETEPLOGCTDEDAVASADPSSMLSEEEKELAEV
QLEDEITLROVLSAKERHVLVEIKQKGLGMNLMNELKONFSKSMHDMOTTAYKKTHET
LSHGQKATAAFSNVGTAKSKFGDMRRK"
join(127425..127443,AL121938.10:41776..41891,
AL121938.10:50816..50964,AL121938.10:69980..70081,
AL121938.10:84532..84541)

Query Match      87.2% Score 21.8; DB 8; Length 151875;
Best Local Similarity 92.0%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGACAAATCT 25
    |||||
Db 111048 AAAAAAAAAATCAACAATCT 111072

RESULT 25
AC019270/C AC019270 154957 bp DNA linear PRI 14-MAR-2002
LOCUS Homo sapiens chromosome 8, clone RP11-10C8, complete sequence.
DEFINITION AC019270
ACCESSION AC019270.12 GI:18767544
VERSION HTG.
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 154957)
Homo sapiens chromosome 8, clone RP11-10C8
Unpublished
2 (bases 1 to 154957)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Bouhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hago,B., Heaford,A., Horton,L.,
Howard,J., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McNeer,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rottman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 154957)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

```

```

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gilde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
Lander,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Mathews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Menus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 154957)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Chapel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
Lander,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Mathews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 20, 2002 this sequence version replaced gi:18252020.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3046
Center clone name: 10_C_8
-----
FEATURES
source
1. 154957
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-10C8"
/clone_1b="RPC1-11 Human Male BAC"
2216..2329
repeat_region

```

```
repeat_region /rpt_family="MER113"
complement(12502..2647)
repeat_region /rpt_family="MIR3"
3423..3731
repeat_region /rpt_family="AluY"
4765..4905
repeat_region /rpt_family="MIR"
5524..5693
repeat_region /rpt_family="(TA)n"
6476..6592
repeat_region /rpt_family="FLAM_C"
6593..6616
repeat_region /rpt_family="(CAT)n"
6617..6634
repeat_region /rpt_family="FLAM_C"
complement(7783..7846)
repeat_region /rpt_family="MER94"
7858..8272
repeat_region /rpt_family="LTR33"
complement(9296..9374)
repeat_region /rpt_family="MIR"
complement(9375..9676)
repeat_region /rpt_family="AluDo"
complement(9677..9870)
repeat_region /rpt_family="MIR"
10733..11113
repeat_region /rpt_family="MSTA"
complement(11165..11453)
repeat_region /rpt_family="AluSg"
11881..12074
repeat_region /rpt_family="MSTB1"
12076..12202
repeat_region /rpt_family="FLAM_C"
12216..12456
repeat_region /rpt_family="MSTB1"
12601..12809
repeat_region /rpt_family="AluSc"
13865..13971
repeat_region /rpt_family="MIR3"
13972..13992
repeat_region /rpt_family="AT_rich"
14119..14422
repeat_region /rpt_family="MT1F"
14458..14583
repeat_region /rpt_family="MT1F"
complement(15103..15138)
repeat_region /rpt_family="MIR"
15630..15932
repeat_region /rpt_family="AluY"
16793..16823
repeat_region /rpt_family="(CA)n"
17216..17279
repeat_region /rpt_family="Tiggr3(Golem)"
17376..17398
repeat_region /rpt_family="AT_rich"
17399..17507
repeat_region /rpt_family="Tiggr3(Golem)"
complement(18015..18184)
repeat_region /rpt_family="MER5A"
complement(18283..18615)
repeat_region /rpt_family="AluY"
18784..18809
repeat_region /rpt_family="AT_rich"
complement(18850..18960)
repeat_region /rpt_family="FLAM_C"
complement(19002..20375)
repeat_region /rpt_family="L2"
complement(21297..21594)
repeat_region /rpt_family="AluDo"
complement(23766..24061)
repeat_region /rpt_family="AluSg"
24391..24412
repeat_region /rpt_family="AT_rich"
```

```
repeat_region complement(24568..24737)
/rpt_family="AluDo"
repeat_region 24738..24767
/rpt_family="(TTTG)n"
repeat_region complement(24768..24891)
/rpt_family="AluDo"
repeat_region complement(25207..26166)
/rpt_family="L2"
repeat_region complement(26170..26373)
/rpt_family="MIR"
repeat_region complement(26384..26635)
/rpt_family="L2"
repeat_region 27277..27325
/rpt_family="AT_rich"
repeat_region 27949..28084
/rpt_family="MIR"
repeat_region 29469..29606
```

Query Match 87.2%; Score 21.8; DB 8; Length 154957;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAATCGACACAAATCT 25
Db 104610 AAAAAAAAAATCGACAAATCT 104586

RESULT 26

AC025367/c

LOCUS Homo sapiens chromosome 3 clone RP11-267D9 map 3, WORKING DRAFT
DEFINITION

SEQUENCE, 26 unordered pieces.

AC025367

AC025367.2 GI:7684485

HTG: HTGS_PHASE1, HTGS_DRAFT.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 155523)

Britten, B., Linton, L., Nussbaum, C. and Lander, E.

Homo sapiens chromosome 3, clone RP11-267D9

2 (bases 1 to 155523)

Unpublished

REFERENCE

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2000 this sequence version replaced gi:7210090.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

```
-----Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----Project Information
Center project name: L7907
Center clone name: 267 D 9
-----Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141691 bases at least Q40
Consensus quality: 148023 bases at least Q30
Consensus quality: 150880 bases at least Q20
Insert size: 155000; agarose-fp
Insert size: 155023; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1073: contig of 1072 bp in length
1073 1172: gap of 100 bp
1173 3339: contig of 2167 bp in length
3340 5115: contig of 176 bp in length
5116 5215: gap of 100 bp
5216 6483: contig of 1268 bp in length
6484 9144: contig of 2561 bp in length
9145 9244: gap of 100 bp
9245 11412: contig of 2168 bp in length
11413 11512: gap of 100 bp
11513 13614: contig of 2102 bp in length
13615 13714: gap of 100 bp
13715 15953: contig of 2239 bp in length
15954 16053: gap of 100 bp
16054 19859: contig of 3806 bp in length
19860 19959: gap of 100 bp
19960 24119: contig of 4160 bp in length
24120 27838: contig of 3619 bp in length
27839 31141: contig of 3203 bp in length
31142 31241: gap of 100 bp
31242 34948: contig of 3707 bp in length
34949 35048: gap of 100 bp
35049 38817: contig of 3769 bp in length
38818 38917: gap of 100 bp
38918 40571: contig of 1654 bp in length
40572 40671: gap of 100 bp
40672 47152: contig of 6481 bp in length
47153 47252: gap of 100 bp
47253 53040: contig of 5788 bp in length
53041 53140: gap of 100 bp
53141 60354: contig of 7214 bp in length
60355 60454: gap of 100 bp
60455 68423: contig of 7469 bp in length
68424 68523: gap of 100 bp
68524 75772: contig of 7249 bp in length
75773 75872: gap of 100 bp
75873 83615: contig of 7743 bp in length
83616 83715: gap of 100 bp
83716 94187: contig of 10472 bp in length
94188 94287: gap of 100 bp
94288 105558: contig of 11271 bp in length
```

```
FEATURES
SOURCE
1. 105559 105658: gap of 100 bp
105659 119578: contig of 13920 bp in length
119579 119679: gap of 100 bp
119679 135655: contig of 15977 bp in length
135656 135755: gap of 100 bp
135756 155523: contig of 19766 bp in length.
Location/Qualifiers
1. 1055523
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-267D9"
/clone_id="RP11 Human Male BAC"
1. 1072
/estimated_length=100
1073. 1172
/estimated_length=100
1173. 3339
/estimated_length=100
3340. 5115
/estimated_length=100
5116. 5215
/estimated_length=100
5216. 6483
/estimated_length=100
6484. 9144
/estimated_length=100
9145. 9244
/estimated_length=100
9245. 11412
/estimated_length=100
11413. 11512
/estimated_length=100
11513. 13614
/estimated_length=100
13615. 13714
/estimated_length=100
13715. 15953
/estimated_length=100
15954. 16053
/estimated_length=100
16054. 19859
/estimated_length=100
19860. 19959
/estimated_length=100
19960. 24119
/estimated_length=100
24120. 27838
/estimated_length=100
27839. 31141
/estimated_length=100
31142. 34948
/estimated_length=100
34949. 35048
/estimated_length=100
35049. 38817
/estimated_length=100
38818. 38917
/estimated_length=100
```

```
misc_feature      38918..40571
                  /note="assembly_fragment
                  clone_end:SP6
                  vector_side:left"
gap              40572..40671
                  /estimated_length=100
misc_feature      40672..47152
                  /note="assembly_fragment"
                  47153..47252
                  /estimated_length=100
gap

Query Match      87.2%; Score 21.8; DB 14; Length 155523;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 AAAAAAAAAATCGAGACCAATCT 25
         |||||
Db      40755 AAAAAAAAAACCGAGACCAATAT 40731

RESULT 27
AC147311      161659 bp      DNA      linear      HTG 07-NOV-2003
LOCUS      Pan troglodytes chromosome UNK clone RP43-51E21, *** SEQUENCING IN
DEFINITION
AC147311      161659 bp      DNA      linear      HTG 07-NOV-2003
AC147311.1.GI:38198851
VERSION      HTG; HTGS_PHASE1.
KEYWORDS      Pan troglodytes (chimpanzee)
SOURCE      Pan troglodytes
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominoidea; Pan.
REFERENCE      1 (bases 1 to 161659)
              Wilson,R.K.
              The sequence of Pan troglodytes clone
              Unpublished
              2 (bases 1 to 161659)
              Wilson,R.K.
              Direct Submision
              Submitted (07-NOV-2003) Genetics, Genome Sequencing Center, 4444
              Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C_PT051E21
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1      1193: contig of 1193 bp in length
*      1194      1293: gap of unknown length
*      1294      2365: contig of 1072 bp in length
*      2366      2465: gap of unknown length
*      2466      4408: contig of 1943 bp in length
*      4409      4508: gap of unknown length
*      4509      5733: contig of 1225 bp in length
*      5734      5833: gap of unknown length
*      5834      7359: contig of 1526 bp in length
*      7360      7459: gap of unknown length
*      7460      10297: contig of 2838 bp in length
*      10298      13797: gap of unknown length
*      13797      13897: contig of 3400 bp in length
*      13898      13897: gap of unknown length
```

```
* 13898      16242: contig of 2345 bp in length
* 16243      16342: gap of unknown length
* 16343      19925: contig of 3583 bp in length
* 19926      20025: gap of unknown length
* 20026      24198: contig of 4173 bp in length
* 24199      24298: gap of unknown length
* 24299      28451: contig of 4153 bp in length
* 28452      28551: gap of unknown length
* 28552      33323: contig of 4681 bp in length
* 33233      33332: gap of unknown length
* 33333      36785: contig of 3453 bp in length
* 36786      36885: gap of unknown length
* 36886      43286: contig of 6400 bp in length
* 43286      43386: gap of unknown length
* 43386      48870: contig of 5485 bp in length
* 48871      48970: gap of unknown length
* 48971      55808: contig of 6838 bp in length
* 55809      55908: gap of unknown length
* 55909      62392: contig of 6484 bp in length
* 62393      62492: gap of unknown length
* 62493      68294: contig of 5802 bp in length
* 68295      68394: gap of unknown length
* 68395      74745: contig of 6351 bp in length
* 74746      74845: gap of unknown length
* 74846      85310: contig of 10465 bp in length
* 85311      85410: gap of unknown length
* 85411      92691: contig of 7281 bp in length
* 92692      92791: gap of unknown length
* 92792      100141: contig of 7350 bp in length
* 100142      100241: gap of unknown length
* 100242      111695: contig of 11454 bp in length
* 111696      111795: gap of unknown length
* 111796      122417: contig of 10622 bp in length
* 122418      122517: gap of unknown length
* 122518      134233: contig of 11716 bp in length
* 134234      134333: gap of unknown length
* 134334      161659: contig of 27326 bp in length.
*
FEATURES
source
1..161659
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    /chromosome="UNK"
    /clone="RP43-51E21"
1..1193
    /note="assembly_name:Contig23"
1194..1293
    /estimated_length=unknown
1294..2365
    /note="assembly_name:Contig30"
2366..2465
    /estimated_length=unknown
2466..4408
    /note="assembly_name:Contig40"
4409..4508
    /estimated_length=unknown
4509..5733
    /note="assembly_name:Contig41"
5734..5833
    /estimated_length=unknown
5834..7359
    /note="assembly_name:Contig43"
7360..7459
    /estimated_length=unknown
7460..10297
    /note="assembly_name:Contig44"
10298..10397
    /estimated_length=unknown
10398..13797
    /note="assembly_name:Contig45"
13798..13897
    /estimated_length=unknown
13898..16242
```

```
gap /note="assembly_name:Contig46"
16243..16342
/misc_feature /estimated_length=unknown
16343..19525
/note="assembly_name:Contig47"
gap 19526..20025
/misc_feature /estimated_length=unknown
20026..24198
/note="assembly_name:Contig48"
gap 24199..24298
/misc_feature /estimated_length=unknown
24299..28451
/note="assembly_name:Contig49"
gap 28452..28551
/misc_feature /estimated_length=unknown
28552..33332
/note="assembly_name:Contig50"
gap 33233..33332
/misc_feature /estimated_length=unknown
33333..36785
/note="assembly_name:Contig51"
gap 36786..36885
/misc_feature /estimated_length=unknown
36886..43385
/note="assembly_name:Contig52"
gap 43286..43385
/misc_feature /estimated_length=unknown
43386..448570
/note="assembly_name:Contig53"
gap 44871..448970
/misc_feature /estimated_length=unknown
448971..55808
/note="assembly_name:Contig54"
gap 55809..55908
/misc_feature /estimated_length=unknown
55909..62392
/note="assembly_name:Contig55"
gap 62393..62492
/misc_feature /estimated_length=unknown
62493..68294
/note="assembly_name:Contig56"
gap 68295..68394
/misc_feature /estimated_length=unknown
68395..74745
/note="assembly_name:Contig57"
gap 74746..74845
/misc_feature /estimated_length=unknown
74846..85310
/note="assembly_name:Contig58"
gap 85311..85410
/misc_feature /estimated_length=unknown
85411..92591
/note="assembly_name:Contig59"
gap 92592..92791
/misc_feature /estimated_length=unknown
92792..100141
/note="assembly_name:Contig60"
gap 100142..100241
/misc_feature /estimated_length=unknown
100242..111695
/note="assembly_name:Contig61"
gap 111696..111795
/misc_feature /estimated_length=unknown
111796..122417
/note="assembly_name:Contig62"
gap 122418..122517
/misc_feature /estimated_length=unknown
122518..134233
/note="assembly_name:Contig63"
gap 134234..134333
/misc_feature /estimated_length=unknown
134334..161659
/note="assembly_name:Contig64"
```

```
ORIGIN
Query Match      87.2%; Score 21.8; DB 14; Length 161659;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1
Db 48080 AAAAAAAAAATCGACGCAAAATCT 25
AAAAAAAAAAAAAAAAATCGCAAAATCT 48104

RESULT 28
AC135279
LOCUS Homo sapiens 12 BAC RP11-114F3 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
ACCESSION AC135279
VERSION AC135279.3 GI:27901816
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 161841)
Muzny,D.M., Adams,C., Adio-Oduola,B., Allosman,F.R., Allen,C.,
Alshrooke,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbarta,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowls,S., Brieva,M., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.J.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhury,I., Christophe,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Decker,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
Escotto,M., Faller,T., Ferraguto,D., Flegny,N., Ford,J., Foster,P.,
Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoque,M.,
Holloway,C., Hollins,B., Homai,F., Howard,S., Huber,J., Hulyk,S.,
Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Lu,X.,
Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Manua,P.,
Marandel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
McWhiney,E., McLeod,M.P., Meador,M., Mei,G., Mettcher,S.,
Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
Mohlbadat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokemko,S., Ogih,M., Okumura,D.,
Oragunye,N., Oviedo,R., Pace,A., Peyton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Pitman,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherrer,S., Scott,G., Shen,H., Shm,C., Shoemaker,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svaltek,A., Taber,P., Tamerias,A., Tamerias,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tuman,K., Vasquez,L., Vera,V., Villanov,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Zuchterlapati,R.,
Weinstock,G. and Gibbs,R.
Direct Submission
TITLE JOURNAL
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 161841)
Worley,K.C.
REFERENCE Direct Submission
JOURNAL Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
```


of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 161841)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 161841)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 5 (bases 1 to 161841)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jan 25, 2003 this sequence version replaced gi:27819428.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source	location/Qualifiers
1..161841	
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="12"
	/clone="RP11-114F3"
	/complement(1..2000)
misc_feature	/note="overlaps bases 1..2000 of clone AC136977"
	/function="clone overlap"
repeat_region	838..862
	/rpt_family="AT_rich"
repeat_region	/complement(936..1150)
	/rpt_family="Aluub"
repeat_region	complement(1639..1650)

/rpt_family="AluSx"
1651..1694
/rpt_family="(TG)n"
complement(1695..1975)
/rpt_family="AluSx"
2005..2217
/rpt_family="MIR"
complement(2398..2616)
/rpt_family="L1MA2"
complement(2629..2798)
/rpt_family="L1MD2"
complement(2815..3141)
/rpt_family="MER7A"
3956..3993
/rpt_family="AT_rich"
complement(4063..4220)
/rpt_family="PAM"
4283..4311
/rpt_family="AT_rich"
4780..4802
/rpt_family="AT_rich"
6237..6261
/rpt_family="AT_rich"
7248..7555
/rpt_family="AluSg"
7867..9243
/rpt_family="L1PA16"
9244..9552
/rpt_family="AluY"
9553..10132
/rpt_family="L1PA16"
10226..10428
/rpt_family="L1PA16"
complement(10554..10864)
/rpt_family="AluUo"
11516..11760
/rpt_family="L2"
complement(11761..12056)
/rpt_family="AluSg"
12057..12403
/rpt_family="L2"
12382..12574
/rpt_family="L2"
complement(12799..13032)
/rpt_family="MER30"
14051..14333
/rpt_family="AluSx"
14356..14539
/rpt_family="AluSg/x"
15160..15531
/rpt_family="MER20B"
15540..15698
/rpt_family="L2"
complement(16385..16424)
/rpt_family="L2"
16969..17173
/rpt_family="L2"
complement(17193..17485)
/rpt_family="AluSg"
complement(17513..17787)
/rpt_family="AluUo"
18942..18964
/rpt_family="AT_rich"
complement(19765..20435)
/rpt_family="L1MB7"
complement(20436..20738)
/rpt_family="AluUo"

Query Match 87.2% Score 21.8; DB 8; length 161841;
Best Local Similarity 92.0%; Pred.No.2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCCGACCAATCT 25

Db 91936 AAAAAAAAAATCCCAAAAAAAAAATCT 91960

|||||

RESULT 29
AC130885/C
LOCUS
DEFINITION Homo sapiens 3 BAC RP11-491D12 (Roswell Park Cancer Institute Human BAC library) complete sequence.
ACCESSION AC130885
VERSION AC130885.3
KEYWORDS GI:22549590
SOURCE HTG.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 163237)
Muzny,D.M., Adams,C., Adio-Odola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayala,M., Banks,T., Barbata,J., Benton,J., Bimage,K., Blanchburg,K., Bonnin,D., Bouck,J., Bowe,S., Bivleva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Din,H.H., Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Matinez,B., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Monabb,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,B., Nwokenkwo,S., Ogun,M., Okwunu,G., Ogunye,N., Ovedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmali,J., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlarczyk,R., Wooder,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163237)
AUTHORS Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (15-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL 3 (bases 1 to 163237)
REFERENCE Direct Submission
AUTHORS Worley,K.C.
JOURNAL Submitted (29-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 163237)
AUTHORS Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (30-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL 5 (bases 1 to 163237)
REFERENCE Direct Submission
AUTHORS Worley,K.C.
JOURNAL Submitted (28-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL 6 (bases 1 to 163237)
REFERENCE Direct Submission
AUTHORS Worley,K.C.
JOURNAL Submitted (15-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL On Aug 30, 2002 this sequence version replaced gi:22538343.
COMMENT INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8086/quality.info/genbank.annotation.html>.

FEATURES
SOURCE Location/Qualifiers
1..163237
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-491D12"
/complement(1..2093)
/note="overlaps bases 1..2091 of clone AC123511"
/function="clone overlap"
21673..21916
/standard_name="SHGC-77130"
22376..22496
/standard_name="RH66762"
26681..27222
/standard_name="G67534"
26827..26933

Query Match	Best Local Similarity	92.0%	Pred. NO. 2.6e+02	Matches	23	Conservative	0	Mismatches	2	Indels	0	Gaps	0
Db	39152	AAAAAAAAAATCGACGACAAATCT	25	AAAAAAAAAATCGACGACAAATAT	39128								
ORIGIN													
Query Match	87.2%	Score 21.8	DB 8	Length 163237									
Best Local Similarity	92.0%	Pred. NO. 2.6e+02											
Matches	23	Conservative	0	Mismatches	2	Indels	0	Gaps	0				
STS		/standard_name="DJS3933"		28065. .28565									
STS		/standard_name="G67533"		29039. .29367									
STS		/standard_name="G67532"		31231. .31626									
STS		/standard_name="G67531"		40356. .40666									
STS		/standard_name="G67530"		43275. .43722									
STS		/standard_name="G67529"		43412. .43573									
STS		/standard_name="RH17368"		57383									
misc_feature		/note="AC130885,C/AC123511.T"		57416. .57427									
misc_feature		/function="polymorphic site"		67128. .67429									
STS		/note="AC130885,12 AS/AC123511.14 AS"		112963. .113181									
STS		/function="polymorphic site"		112963. .113181									
STS		/standard_name="G59369"		112963. .113181									
STS		/standard_name="D1S3059"											
ORIGIN													
Query Match	87.2%	Score 21.8	DB 8	Length 163237									
Best Local Similarity	92.0%	Pred. NO. 2.6e+02											
Matches	23	Conservative	0	Mismatches	2	Indels	0	Gaps	0				
Db	39152	AAAAAAAAAATCGACGACAAATCT	25	AAAAAAAAAATCGACGACAAATAT	39128								
ORIGIN													
Query Match	87.2%	Score 21.8	DB 8	Length 163237									
Best Local Similarity	92.0%	Pred. NO. 2.6e+02											
Matches	23	Conservative	0	Mismatches	2	Indels	0	Gaps	0				
Db	39152	AAAAAAAAAATCGACGACAAATCT	25	AAAAAAAAAATCGACGACAAATAT	39128								
ORIGIN													
Query Match	87.2%	Score 21.8	DB 8	Length 163237									
Best Local Similarity	92.0%	Pred. NO. 2.6e+02											
Matches	23	Conservative	0	Mismatches	2	Indels	0	Gaps	0				
Db	39152	AAAAAAAAAATCGACGACAAATCT	25	AAAAAAAAAATCGACGACAAATAT	39128								
ORIGIN													
Query Match	87.2%	Score 21.8	DB 8	Length 163237									
Best Local Similarity	92.0%	Pred. NO. 2.6e+02											
Matches	23	Conservative	0	Mismatches	2	Indels	0	Gaps	0				
Db	39152	AAAAAAAAAATCGACGACAAATCT	25	AAAAAAAAAATCGACGACAAATAT	39128								
ORIGIN													
Query Match	87.2%	Score 21.											

```

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4019
Center clone name: 21 H 8
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158445 bases at least Q40
Consensus quality: 161965 bases at least Q40
Consensus quality: 161997 bases at least Q20
Insert size: 167000; agarose-fp
Quality coverage: 5.3 in Q20 bases; agarose-fp
Quality coverage: 5.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1      1174: contig of 1174 bp in length
* *
* 1175   1274: gap of 100 bp
* *
* 1275   4788: contig of 3514 bp in length
* *
* 4789   4888: gap of 100 bp
* *
* 4889   14973: contig of 9985 bp in length
* *
* 14974   14973: gap of 100 bp
* *
* 14974   32556: contig of 17583 bp in length
* *
* 32557   32557: gap of 100 bp
* *
* 32557   52576: contig of 19920 bp in length
* *
* 52577   52576: gap of 100 bp
* *
* 52677   73204: contig of 20528 bp in length
* *
* 73205   73304: gap of 100 bp
* *
* 73305   109834: contig of 36530 bp in length
* *
* 109835   109934: gap of 100 bp
* *
* 109935   163554: contig of 53620 bp in length.
*
Location/Qualifiers
1. .163554
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /clone="RP11-21H8"
    /clone_id="RPCI-11 Human Male BAC"
    /clone_1174
1. .1174
    /note="assembly_fragment"
1175..1274
    /estimated_length=100
1275..4788
    /note="assembly_fragment"
4789..4888
    /estimated_length=100
4889..14873
    /note="assembly_fragment"
14874..14973
    /estimated_length=100
14974..32556
    /note="assembly_fragment"
32557..32556
    /estimated_length=100
32561..52576
    /note="assembly_fragment"
vector side:left"

```

```

gap
52577..52676
/estimated_length=100
misc_feature
52677..73304
/note="assembly_fragment"
gap
73205..73304
/estimated_length=100
misc_feature
73305..109834
/note="assembly_fragment"
gap
109835..109934
/estimated_length=100
misc_feature
109935..163554
/note="assembly_fragment"
clone_end=77
vector_side:right"

ORIGIN

Query Match      87.2%; Score 21.8; DB 14; Length 163554;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 9454 AAAAAAAAAATCGCAGCAAAATCT 9478

RESULT 31
AC008949 163731 bp DNA linear PRI 23-AUG-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2335024, complete sequence.
ACCESSION AC008949
VERSION AC008949.9 GI:15281188
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 163731)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163731)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163731)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Aug 23, 2001 this sequence version replaced gi:13699462.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sbgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.5.

FEATURES
Source
1..163731
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2335024"

ORIGIN

Query Match      87.2%; Score 21.8; DB 8; Length 163731;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGCAGCAAAATCT 25

```

```

Db 61421 AAAAAAAAAATCGCAGCAAAATCT 61397

RESULT 32
AL591663/c 166517 bp DNA linear HTG 19-DEC-2001
LOCUS Homo sapiens chromosome 10 clone RPJ3-206M19, WORKING DRAFT
DEFINITION SEQUENCE, 28 unordered pieces.
ACCESSION AL591663.2 GI:14529990
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.

1
Burton, J.
Direct Submission
Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 22, 2001 this sequence version replaced gi:14268334.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: hb206M19
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 150526 bases at least Q40
Consensus quality: 156647 bases at least Q30
Consensus quality: 160260 bases at least Q20
Insert size: 163817; sum-of-contigs
Insert size: 256494; 24.1% error; agarose-fp
Quality coverage: 3.02x in Q20 bases; sum-of-contigs Quality
coverage: 2.23x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6863: contig of 6863 bp in length
6864 6863: gap of 100 bp
6864 14905: contig of 7942 bp in length
6864 15005: gap of 100 bp
6864 17487: contig of 2482 bp in length
6864 17587: gap of 100 bp
6864 23573: contig of 5986 bp in length
6864 23673: gap of 100 bp
6864 32984: contig of 9311 bp in length
6864 33085: gap of 100 bp
6864 37287: contig of 4203 bp in length
6864 37387: gap of 100 bp
6864 43136: contig of 5749 bp in length
6864 43236: gap of 100 bp
6864 43237: gap of 100 bp
6864 47819: contig of 4583 bp in length
6864 47919: gap of 100 bp
6864 47920: contig of 2772 bp in length
6864 50691: gap of 100 bp
6864 50792: contig of 6219 bp in length
6864 57110: gap of 100 bp
6864 57111: contig of 3712 bp in length
6864 60922: gap of 100 bp
6864 60923: contig of 2933 bp in length

```

```
* 63856 63955: gap of 100 bp
* 63956 74630: contig of 10675 bp in length
* 74630 74730: gap of 100 bp
* 74731 83250: contig of 8520 bp in length
* 83250 83350: gap of 100 bp
* 83351 93216: contig of 9866 bp in length
* 93217 93316: gap of 100 bp
* 93317 98003: contig of 4687 bp in length
* 98004 98103: gap of 100 bp
* 98104 107042: contig of 8939 bp in length
* 107043 107143: gap of 100 bp
* 107143 111209: contig of 4057 bp in length
* 111210 111309: gap of 100 bp
* 111310 117200: contig of 5891 bp in length
* 117201 117300: gap of 100 bp
* 117301 121070: contig of 3770 bp in length
* 121071 121170: gap of 100 bp
* 121171 128898: contig of 7728 bp in length
* 128899 128998: gap of 100 bp
* 128999 131416: contig of 2418 bp in length
* 131417 131516: gap of 100 bp
* 131517 137669: contig of 6153 bp in length
* 137670 137769: gap of 100 bp
* 137770 145371: contig of 7602 bp in length
* 145372 145471: gap of 100 bp
* 145472 148522: contig of 3051 bp in length
* 148523 148622: gap of 100 bp
* 148623 151178: contig of 2556 bp in length
* 151179 151278: gap of 100 bp
* 151279 162775: contig of 11497 bp in length
* 162776 162875: gap of 100 bp
* 162876 16517: contig of 3642 bp in length.
```

FEATURES

source

```
1..16517
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-206M19"
/clone_1ib="RPC1-13.1"
1..6863
/note="assembly_fragment:00213
fragment_chain:1"
misc_feature
6964..14905
/note="assembly_fragment:01337
fragment_chain:1"
misc_feature
15006..17487
/note="assembly_fragment:00737
fragment_chain:1"
misc_feature
17588..23573
/note="assembly_fragment:01703
fragment_chain:1"
misc_feature
23674..32984
/note="assembly_fragment:01082
fragment_chain:1"
misc_feature
33085..37287
/note="assembly_fragment:01089
fragment_chain:1"
misc_feature
37388..43136
/note="assembly_fragment:01752
fragment_chain:1"
misc_feature
43237..47819
/note="assembly_fragment:00709
fragment_chain:1"
misc_feature
47920..50691
/note="assembly_fragment:01258
fragment_chain:1"
misc_feature
50792..57010
/note="assembly_fragment:01159
fragment_chain:2"
misc_feature
57111..60822
/note="assembly_fragment:01165
fragment_chain:2"
```

```
misc_feature 60923..63855
/note="assembly_fragment:01793
fragment_chain:2"
misc_feature 63956..74630
/note="assembly_fragment:01144
fragment_chain:2"
misc_feature 74731..83250
/note="assembly_fragment:00135
fragment_chain:3"
misc_feature 83351..93216
/note="assembly_fragment:01427
fragment_chain:3"
misc_feature 93317..98003
/note="assembly_fragment:00320
fragment_chain:3"
misc_feature 98104..107042
/note="assembly_fragment:01115
fragment_chain:4"
misc_feature 107143..111209
/note="assembly_fragment:00036
fragment_chain:4"
misc_feature 111310..117200
/note="assembly_fragment:01713
fragment_chain:4"
misc_feature 117301..121070
/note="assembly_fragment:00152"
121171..128898
/note="assembly_fragment:00190"
128899..131416
/note="assembly_fragment:00362"
131517..137669
/note="assembly_fragment:00390"
137770..145371
/note="assembly_fragment:00669"
145472..148522
/note="assembly_fragment:00694"
148623..151178
/note="assembly_fragment:00793"
151279..162775
/note="assembly_fragment:00836"
162876..16517
/note="assembly_fragment:01551"
misc_feature
```

ORIGIN

Query Match

Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGACACAAATCT 25

DB 138850 AAAAAAAAAATCGACACAAATCT 138826

RESULT 33

AC021706/c AC021706 168833 bp DNA linear HTG 08-APR-2000

LOCUS Homo sapiens clone RP11-388017, WORKING DRAFT SEQUENCE, 33

DEFINITION unordered pieces.

ACCESSION AC021706

VERSION AC021706.3 GI:7528116

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 168833)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens, clone Rpl1-388017

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 168833)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

TITLE
JOURNAL
COMMENT

Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Darellano, K., Dewar, K., Domino, M., Doyle, M., Fensholt, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Garday, S., Grant, G., Hago, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lander, T., Lechocz, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Mackdonald, P., Margulis, N., McEwan, P., McCurk, A., McKernan, K.,
McSheeters, R., Meldrim, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testave, S., Theodore, J.,
Tirrell, A., Vassiliou, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

Direct Submission
Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced gi:6893688.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
Project Information

Center project name: 388_O_17
Center clone name: 15882

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 149025 bases at least Q40
Consensus quality: 158058 bases at least Q30

Consensus quality: 162087 bases at least Q20
Insert size: 16600; agarose-fp

Insert size: 165633; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp

Quality coverage: 3.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 33 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be replaced.

1 1124: contig of 1124 bp in length
* 1125 1224: gap of 100 bp
* 1225 1224: contig of 1022 bp in length
* 1226 2346: gap of 100 bp
* 2347 3895: contig of 1549 bp in length
* 3896 3995: gap of 100 bp
* 3996 5068: contig of 1073 bp in length
* 5069 5168: gap of 100 bp
* 5169 6557: contig of 1389 bp in length
* 6558 6657: gap of 100 bp
* 6659 8301: contig of 1644 bp in length
* 8302 8402: gap of 100 bp
* 8403 9758: contig of 1357 bp in length
* 9759 9858: gap of 100 bp
* 9859 12176: contig of 2318 bp in length
* 12177 12276: gap of 100 bp
* 12277 14646: contig of 2370 bp in length
* 14647 14746: gap of 100 bp
* 14747 17380: contig of 2634 bp in length
* 17381 17480: gap of 100 bp
* 17481 19793: contig of 2313 bp in length
* 19794 19893: gap of 100 bp
* 19894 22375: contig of 2482 bp in length

FEATURES	source
misc_feature	1. 1124 /note="assembly_fragment"
gap	1125. 1224 /estimated_length=100
misc_feature	1225. 2246 /note="assembly_fragment"
gap	2247. 2346 /estimated_length=100
misc_feature	2347. 3895 /note="assembly_fragment"
gap	3896. 3995 /estimated_length=100
misc_feature	3996. 5068 /note="assembly_fragment"
gap	5069. 5168 /estimated_length=100
misc_feature	5169. 6557 /note="assembly_fragment"
gap	6558. 6657 /estimated_length=100
misc_feature	6659. 8301 /note="assembly_fragment"
gap	8302. 8401 /estimated_length=100

```

misc_feature      8402..9758
                  /note="assembly_fragment"
gap              9759..9858
                  /estimated_length=100
misc_feature      9859..12176
                  /note="assembly_fragment"
gap              12177..12276
                  /estimated_length=100
misc_feature      12277..14646
                  /note="assembly_fragment"
gap              14647..14746
                  /estimated_length=100
misc_feature      14747..17380
                  /note="assembly_fragment"
gap              17381..17480
                  /estimated_length=100
misc_feature      17481..19793
                  /note="assembly_fragment"
gap              19794..19893
                  /estimated_length=100
misc_feature      19894..22375
                  /note="assembly_fragment"
gap              22376..22475
                  /estimated_length=100
misc_feature      22476..24401
                  /note="assembly_fragment"
                  clone_end=17
gap              24402..24501
                  /estimated_length=100
misc_feature      24502..27437
                  /note="assembly_fragment"
gap              27438..27537
                  /estimated_length=100
misc_feature      27538..29420
                  /note="assembly_fragment"

Query Match      87.2%; Score 21.8; DB 14; Length 168833;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAATCT 25
Db 6904 AAAAAAAAAATCGCAAAAAATCT 6880

RESULT 34
AL353616
LOCUS             AL353616          169993 bp          DNA          linear          PRI 18-MAY-2005
DEFINITION        Human DNA sequence from clone RP11-274B18 on chromosome 9 Contains
                   the 3' end of the PGM5 gene for phosphoglucomutase 5 (PGMP), a
                   novel gene (MGC34760), a squamous cell carcinoma antigen recognized
                   by T cell (SART2) pseudogene and 2 novel genes, complete sequence.
ACCESSION          AL353616
VERSION            AL353616.14  GI:18250740
KEYWORDS            HTG; MGC34760; PGM5; PGMRP; SART2.
SOURCE             Homo sapiens
                   Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                   Homnidae; Homo.
                   1 (bases 1 to 169993)
REFERENCE          1. Leongamornlert, D.
AUTHORS            Direct Submission
JOURNAL            Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
                   Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
COMMENT            Clone request: clonerequest@sanger.ac.uk
                   On Jan 19, 2002 this sequence version replaced gi:12584384.
                   The following abbreviations are used to associate primary accession
                   numbers given in the feature table with their source databases:
                   Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
                   on the WORMPEP database can be found at
                   http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

```

```

FEATURES
source
1..169993
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="9"
   /clone="RP11-274B18"
   /clone_lib="RPCI-11.1"
1
   /note="Clone left end: RP11-274B18"
   join(AL353608..20:136408..171921,AL161457..13:73550..107818,
15636..17130)
   /gene="PGM5"
   /locus_tag="RP11-274B18.5-001"
   join(AL353608..20:136408..136836
AL353608..20:157647..157809,AL353608..20:163846..163992,
AL353608..20:166911..167036,AL353608..20:170982..171172,
AL353608..20:171767..171921,AL161457..13:73550..73665,
AL161457..13:87875..88010,AL161457..13:92322..92505,
AL161457..13:107684..107818,15636..17130)
   /gene="PGMS"
   /locus_tag="RP11-274B18.5-001"
   /product="phosphoglucomutase 5"
   /note="match: ESTs: AA631141.1 AA706788.1 AL046941.1
AV596258.1 AV723073.1 AW652952.1 AW652958.1 BB264139.2
BB621396.1 BB628028.1 BB641261.1 BE221369.1 BE235707.1
BF058433.1 BF060711.1 BF230880.1 BF306296.1 BF306987.1
BG254110.1 B1346230.1 B1847349.1 BM363295.1 BM363453.1
BM364512.1 BM366644.1 BM699018.1 BM975035.1 BQ001679.1
BQ007078.1 BQ011705.1 BQ011902.1 R86043.1
match: cDNAs: L40933"
   join(AL353608..20:136759..136836
AL353608..20:157647..157809,AL353608..20:163846..163992,
AL353608..20:166911..167036,AL353608..20:170982..171172,
AL353608..20:171767..171921,AL161457..13:73550..73665,
AL161457..13:87875..88010,AL161457..13:92322..92505,
AL161457..13:107684..107818,15636..15725)
   /gene="PGMS"
   /locus_tag="RP11-274B18.5-001"
   /standard_name="OTTHUMP00000021418"
   /note="match: proteins: Q15124"
   /codon_start=1
   /product="phosphoglucomutase 5"
   /protein_id="CAH71906.1"
   /db_xref="GI:55652858"
   /db_xref="InterPro:IPR005841"
   /db_xref="InterPro:IPR005843"
   /db_xref="InterPro:IPR005844"
   /db_xref="InterPro:IPR005845"
   /db_xref="InterPro:IPR005846"
   /translation="MVGSGDGRVSRPRRIEIVYQMAANGIGRIIIGNGIISLPAYS
CIIRKIAAGIILITASHCPGSGEGVKNFNVANGAPADPVDSKTIQISKTIEEYA

```

was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>

RP11-274B18 is from the library RPCI-11.1 constructed by the group of Pictet de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

Location/Qualifiers

```

misc_feature
/locus_tag="RP11-274B18.4-001"
/note="Clone_left_end: RP11-203L2"
169993
/note="Clone_right_end: RP11-274B18"

ORIGIN

Query Match      87.24; Score 21.8; DB 8; Length 169993;
Best Local Similarity 92.04; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY      1 AAAAAAAAAATCGCAGCAATCT 25
|||||
Db      81899 AAAAAAAAAATCGCAAAAAATCT 81923

RESULT 35
AC022844
LOCUS      AC022844      171732 bp      DNA      linear      PRI 22-NOV-2001
DEFINITION Homo sapiens chromosome 8, clone RP11-250P18, complete sequence.
ACCESSION  AC022844
VERSION     AC022844.8  GI:17048275
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eumalia; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 171732)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-250P18
Unpublished
2 (bases 1 to 171732)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,Y., Bede,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fenebor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
McNeeters,R., Meldrim,J., Menais,P., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rochman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wymann,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 171732)
Biren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Baatien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Glasde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labrecque,K.,
Lamaszere,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLenn,C., Macdonald,P., Major,J., Margulis,N., Mathews,C.,
McCarthy,L., McEwan,P., McKernan,K., McNeeters,R., Meldrim,J.,
Meyers,H., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Nobu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhng,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rossetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J.,
Tophan,K., Travers,M., Travis,N., Triticio,J., Vassiliev,H.,

```


TITLE
JOURNAL
COMMENT

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research: 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2001 this sequence version replaced gi:16118193.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www-seg.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L6161

Center clone name: 250_P_18

FEATURES

source

Location/Qualifiers
1..171732
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-250P18"
/clone_lib="RPC1-11 Human Male BAC"
complement(306..380)
/rpt_family="MSTC"
404..433
/rpt_family="AT_rich"
complement(1195..1281)
/rpt_family="Zaphod"
3566..3597
/rpt_family="AT_rich"
complement(3799..3904)
/rpt_family="L2"
4689..4729
/rpt_family="TG)n"
4846..4914
/rpt_family="T-rich"
complement(4916..5427)
/rpt_family="L1PB2"
5429..5520
/rpt_family="L1PB2"
5728..5766
/rpt_family="AT_rich"
5852..5982
/rpt_family="L1ME1"
5986..7066
/rpt_family="L1ME1"
7067..7361
/rpt_family="AluY"
7362..7821
/rpt_family="L1ME1"
7811..7876
/rpt_family="L1ME1"
9138..9174
/rpt_family="(CA)n"
9419..9700
/rpt_family="AluY"
9701..9741
/rpt_family="(CAT)n"
10129..10280
/rpt_family="MIR"
10862..10958
/rpt_family="MIR"
10962..11266
/rpt_family="AluY"
12275..12434
/rpt_family="MIR"
complement(12795..12886)
/rpt_family="SS"

repeat_region 13221..13245
/rpt_family="(T)n"
repeat_region 13602..13907
/rpt_family="L2"
complement(14878..15223)
/rpt_family="THE1C"
complement(15449..15541)
/rpt_family="MIR3"
15940..15969
/rpt_family="AT_rich"
16167..16309
/rpt_family="FRAM"
16317..16347
/rpt_family="(GA)n"
16379..16481
/rpt_family="MIR"
complement(16625..16791)
/rpt_family="AluDb"
complement(17312..17609)
/rpt_family="AluSx"
complement(17660..18024)
/rpt_family="MLT1A2"
19646..19689
/rpt_family="L2"
20540..20590
/rpt_family="AT_rich"
21152..21204
/rpt_family="MIR"
21222..21367
/rpt_family="L1ME1"
21419..21654
/rpt_family="L1ME1"
complement(21658..21736)
/rpt_family="MER20"
21808..21983
/rpt_family="L1ME1"
complement(21984..22337)
/rpt_family="L1PA5"
22338..24087
/rpt_family="L1ME1"
24488..24806
/rpt_family="AluSx"
24904..25311
/rpt_family="L2"
25295..26234
/rpt_family="L2"
/rpt_family="L2"
complement(26235..26570)
/rpt_family="AluSg"
26571..26631
/rpt_family="L2"
complement(27293..27454)
/rpt_family="L2"
complement(27494..27913)
/rpt_family="L1PA9"
complement(28281..29147)
/rpt_family="L1PA7"
29488..29768
/rpt_family="AluY"
29769..29917
/rpt_family="L1PA5"
29919..29959
/rpt_family="AT_rich"
30507..30819
/rpt_family="AluSx"
31981..32158
/rpt_family="MIR"
33793..33818
/rpt_family="(TTTC)n"
complement(35648..35843)
/rpt_family="MIR3"
35924..35948

Query Match

87.2%; Score 21.8; DB 8; Length 171732;

Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACCAATCT 25
|||||
Db 135577 AAAAAAAAAATCGCAAAAATCT.135601

RESULT 36

AL451107

LOCUS

AL451107 172779 bp DNA linear PRI 18-MAY-2005
Human DNA sequence from clone RP11-14C22 on chromosome 10 contains

the gene for a novel protein similar to lysozyme C-1 (1
4-beta-N-acetylmuramidase C EC 3.2.1.17) (MG033408), three novel
genes and a superovillin pseudogene (SVIL), complete sequence.

ACCESSION AL451107 AC025346

AL451107.6 GI:14148855

HTG; lysozyme; MG033408; superovillin; SVIL.

KEYWORDS Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.

1 (bases 1 to 172779)

REFERENCE

Dunn, M.

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone request: clonerequest@sanger.ac.uk
On May 18, 2001 this sequence version replaced gi:13992064.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi., EMBL; Sw., SWISSPROT; Tr., TREMBL; Mp., WormPep; Information
on the WormPep database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10

RP11-14C22 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

Drift Sequence Produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.

FEATURES

SOURCE

1..172779

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-14C22"

/clone_1fb="RPCI-11.1"

misc_feature

1

gene

/note="Clone left end: RP11-14C22"
Complement(join(9406..9483,10605..10718,13161..13238,
15097..15207,16005..16044))
/gene="LYZL2"

mRNA

/locus_tag="RP11-14C22.2-001"
Complement(join(9406..9483,10605..10718,13161..13238,
15097..15207,16005..16044))
/gene="LYZL2"

/locus_tag="RP11-14C22.2-001"
/product="lysozyme-like 2"

/note="match: ESTs: A1671318.1"
Complement(14962)

polyA_site

/gene="LYZL2"

polyA_site

/locus_tag="RP11-14C22.2-001"
Complement(14969)

gene

/gene="LYZL2"

gene

/locus_tag="RP11-14C22.2-001"
Complement(join(14972..15207,16005..16083,29288..29446,
29960..30123,32776..32945))
/gene="LYZL2"

mRNA

/locus_tag="RP11-14C22.2-002"
Complement(join(14972..15207,16005..16083,29288..29446,
29960..30123,32776..32945))
/gene="LYZL2"

mRNA

/locus_tag="RP11-14C22.2-002"
/product="lysozyme-like 2"

mRNA

/note="match: ESTs: A1962816.1 AW614024.1 AW631146.1
BE501011.1 BF062352.1 BF979319.1 BI825941.1 BU568794.1
match: CDNA: AK06241.1 AK06630.1 BC021730.1"
Complement(14972)

polyA_site

/gene="LYZL2"

polyA_site

/locus_tag="RP11-14C22.2-001"
Complement(14990..14995)

polyA_site

/gene="LYZL2"

polyA_site

/locus_tag="RP11-14C22.2-001"
Complement(join(15138..15207,16005..16083,29288..29446,
29960..30123,32776..32888))
/gene="LYZL2"

polyA_site

/locus_tag="RP11-14C22.2-002"
/standard_name="OTTHUMP0000019400"

CDS

/note="match: proteins: P00707 Q8WW16 Q96LF2 Q9CPX3 Q9H1R8
Q9H1R9"

polyA_site

/codon_start=1
/product="lysozyme-like 2"

polyA_site

/protein_id="CAH70607.1"

polyA_site

/db_xref="GI:55664937"

polyA_site

/db_xref="GOA:Q6NZ69"

polyA_site

/db_xref="InterPro:IPR000974"

polyA_site

/db_xref="UniProt/TREMBL:Q6NZ69"

polyA_site

/translation="MDAPLSCSPYTWSSVSADSTKASASAAAGTNPDPFCRLQDA
LRMKAAGILTLIGCLVTGASKITYRCKLAKIFSRAGLDYWGFSLGNIWICMAYEESG
YNTTAQTVLDDGSDIDYGIPOINSFAMCRCKLKENNHCHVACSLVTDLTDIICAK
KIVETQGMNVWQMKKGCRDLSDMKXCEVS"

polyA_site

/locus_tag="RP11-14C22.5-001"

polyA_site

/locus_tag="RP11-14C22.5-001"

polyA_site

/locus_tag="RP11-14C22.5-001"

polyA_site

/product="putative novel transcript"

polyA_site

/note="match: ESTs: AW085317.1"

polyA_site

33224..33229

polyA_site

/locus_tag="RP11-14C22.5-001"

polyA_site

33246

polyA_site

/locus_tag="RP11-14C22.5-001"

polyA_site

33587

polyA_site

/note="Clone right end: RP11-253D19"

polyA_site

/locus_tag="58894.59752.59808.60513..>60718)

polyA_site

/locus_tag="RP11-14C22.3-001"

polyA_site

/locus_tag="58894.59752.59808.60513..>60718)

polyA_site

/locus_tag="RP11-14C22.3-001"

polyA_site

/product="putative novel transcript"

polyA_site

/note="match: ESTs: A1142493.1"

polyA_site

join(74749..74839,88445..88592,92422..92511,92598..92684,
94257..94434,96679..96851,99105..99260,100559..100693,
101300..101424,101553..101698,103270..103445,
106705..106849,107596..107872,112481..112636,
114545..114685,115741..115866,120110..120265,

```

120366..120449)
/locus_tag="RP11-14C22.4-001"
/pseudo
join(74749..74839,88445..88592,92422..92511,92598..92684,
94257..94434,96679..96851,99105..99260,100559..100695,
101300..101424,101553..101698,103270..103445,
106705..106849,107596..107872,112481..112536,
114545..114685,115741..115866,120110..120265,
120366..120449)
/locus_tag="RP11-14C22.4-001"
/notes="match: proteins: O6385 O60611 O60612 O95425 O8K4L2
O8K4L3 O9H1R7"
/pseudo
/codon_start=1
/product="supervillin (SVIL) pseudogene"
join(126434..126519,126602..126923)
/locus_tag="RP11-14C22.6-001"
join(126434..126519,126602..126923)
/locus_tag="RP11-14C22.6-001"
/product="putative novel transcript"
/notes="match: ESTs: A1674082.1"
126898..126903
/locus_tag="RP11-14C22.6-001"
126923
/locus_tag="RP11-14C22.6-001"
15081
/notes="Clone_left_end: RP11-330011"
172779
/notes="Clone_right_end: RP11-14C22"

ORIGIN

Query Match      87.2%; Score 21.8; DB 8; Length 172779;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGAGACAATCT 25
Db 13307 AAAAAAAAAATCGACAATCT 13331

RESULT 37
AC073894
LOCUS
DEFINITION
AC073894
VERSION
AC073894.15 GI:20335601
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 173491)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbista,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doughwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Ehrhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Hawlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homel,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,O., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgied,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mahpey,E., McLeod,M.P., Mendiratta,A., Martinez,B.,
Masey,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokoko,S., Ogul,M., Okunolu,G.,
Ogunyemi,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Pimus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoohat,N., Sison,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Sytek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vaequez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 173491)
Worley,K.C.
Direct Submission
Submitted (01-JUL-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 173491)
Worley,K.C.
Direct Submission
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:18449546.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAZB
Center clone name: RP11-241019
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-terminator Big Dye 5% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 164781 bases at least Q40
Consensus quality: 168264 bases at least Q30
Consensus quality: 169842 bases at least Q20
Estimated insert size: 173762; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3720 3719: contig of 3719 bp in length
3819: gap of unknown length
3820 7606: contig of 3787 bp in length
7607 7707: gap of unknown length
7707 10227: contig of 2521 bp in length
10228 10327: gap of unknown length
10328 17034: contig of 6707 bp in length
17035 17134: gap of unknown length

```



```
source
1..175134
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /cd_xref="taxon:9606"
    /chromosome="9"
    /clone_id="RP11-1K21"
    /clone_lib="RPCI-11.1"
misc_feature
1..3511
    /note="assembly_fragment:00168
    fragment_chain:1"
misc_feature
3512..25895
    /note="assembly_fragment:00194
    fragment_chain:1"
misc_feature
25896..31543
    /note="assembly_fragment:00852
    fragment_chain:2"
misc_feature
31644..39248
    /note="assembly_fragment:00602
    fragment_chain:2"
misc_feature
39349..79948
    /note="assembly_fragment:01553
    fragment_chain:3"
misc_feature
50049..66131
    /note="assembly_fragment:01966
    fragment_chain:3"
misc_feature
66232..80131
    /note="assembly_fragment:02033
    fragment_chain:4"
misc_feature
80232..89785
    /note="assembly_fragment:00372
    fragment_chain:4"
misc_feature
89886..52978
    /note="assembly_fragment:02305
    fragment_chain:5"
misc_feature
93079..97528
    /note="assembly_fragment:00011
    fragment_chain:5"
misc_feature
97629..101520
    /note="assembly_fragment:00955"
    fragment_chain:5"
misc_feature
101621..112966
    /note="assembly_fragment:01248"
    fragment_chain:5"
misc_feature
113067..119217
    /note="assembly_fragment:01337"
    fragment_chain:5"
misc_feature
119318..121556
    /note="assembly_fragment:01523"
    fragment_chain:5"
misc_feature
121657..134400
    /note="assembly_fragment:01568"
    fragment_chain:5"
misc_feature
134501..144934
    /note="assembly_fragment:01617"
    fragment_chain:5"
misc_feature
145035..148818
    /note="assembly_fragment:01636"
    fragment_chain:5"
misc_feature
148919..152178
    /note="assembly_fragment:01972"
    fragment_chain:5"
misc_feature
152279..165020
    /note="assembly_fragment:00951
    fragment_chain:6"
misc_feature
165121..175134
    /note="assembly_fragment:01119
    fragment_chain:6"
vector_end:T7
vector_side:right
```

ORIGIN

Query Match	87.2%;	Score 21.8;	DB 14;	Length 175134;
Best Local Similarity	92.0%;	Pred. No. 2.6e+02;		
Matches 23;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

```
Qy      1 AAAAAAAAAATCGACAAATCT 25
          |||||
Db      26186 AAAAAAAAAATTCAGAAAAATCT 26210
          |||||
```

RESULT 39

AC152354	AC152354	189791 bp	DNA	linear	HTG 03-NOV-2004
LOCUS	Pan troglodytes chromosome X clone RP43-011D19 map human ortholog				
DEFINITION	q26.2, 2 ordered pieces.				
ACCESSION	AC152354				
VERSION	AC152354.1	GI:55251387			
KEYWORDS	HTG PHASE2; HTGS_FULITOP; HTGS_CANCELLED.				
SOURCE	Pan troglodytes (chimpanzee)				
ORGANISM	Pan troglodytes				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.				
AUTHORS	1 (bases 1 to 189791)				
TITLE	Tauden,S., Tazner,S., Baumgart,C., Mueller,O. and Platzer,M.				
JOURNAL	DNA sequence of chimpanzee chromosome X				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 189791)				
TITLE	Lagemann,D. and Platzer,M.				
JOURNAL	Direct Submission				
COMMENT	Submitted (03-NOV-2004) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany				

	Genome Center				

Sequencing clone based: pUC18, 86% of reads
Whole genome shotgun: 14% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Pirap; version 0.990329
Consensus quality: 189654 bases at least Q40
Consensus quality: 189667 bases at least Q30
Consensus quality: 189651 bases at least Q20
Quality coverage: 10.15x

This sequence was generated with the support of external whole genome shotgun data (<http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi>) as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Neighboring sequence information:
This clone is overlapped by RP43-057B02.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 58947: contig of 58947 bp in length
58948 58947: gap of unknown length

```

FEATURES          *      59048      189791: contig of 130744 bp in length.
SOURCE
    1.189791
        Location/Qualifiers
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
            /db_xref="taxon:9598"
            /chromosome="X"
            /map="human ortholog q26.2"
            /clone="RP43-011D19"
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
            /db_xref="taxon:9598"
            /clone="RP43-057802"
            /note="overlapping clone"
    1.43
        /note="single stranded/single chemistry region"
    143..214
        /note="single stranded/single chemistry region"
    2451..2862
        /note="single stranded/single chemistry region"
    2863..2960
        /note="single clone coverage"
    2961..3245
        /note="single stranded/single chemistry region"
    4859..5057
        /note="single stranded/single chemistry region"
    4949..4950
        /note="low quality region"
    15793..15878
        /note="single stranded/single chemistry region"
    21939..21982
        /note="single stranded/single chemistry region"
    23573..23746
        /note="single stranded/single chemistry region"
    26636..26703
        /note="single stranded/single chemistry region"
    28360..28437
        /note="single stranded/single chemistry region"
    29168..29694
        /note="single stranded/single chemistry region"
    42239..42910
        /note="single stranded/single chemistry region"
    43030..43035
        /note="single stranded/single chemistry region"
    43036..43073
        /note="single clone coverage"
    43074..43951
        /note="single stranded/single chemistry region"
    55742..56092
        /note="single stranded/single chemistry region"
    56290..56593
        /note="single stranded/single chemistry region"
    58752..58773
        /note="single stranded/single chemistry region"
    58830..59083
        /note="single stranded/single chemistry region"
    58948..59047
        /estimated_length=unknown
    5948
        /note="low quality region"
    59808..59861
        /note="single stranded/single chemistry region"
    87447..87851
        /note="single stranded/single chemistry region"
    88748..88842
        /note="single stranded/single chemistry region"
    111941..112084
        /note="single stranded/single chemistry region"
    120720..120814
        /note="single stranded/single chemistry region"
    134618..134658
        /note="single clone coverage"

```

```

misc_feature      134618
                  /note="low quality region"
misc_feature      134724..134831
                  /note="single stranded/single chemistry region"
misc_feature      137592..137850
                  /note="single stranded/single chemistry region"
misc_feature      140932..141004
                  /note="whole-genome shotgun only"
misc_feature      144029..144060
                  /note="single stranded/single chemistry region"
misc_feature      16048..166217
                  /note="single stranded/single chemistry region"

```

ORIGIN

```

Query Match      87.2%; Score 21.8; DB 14; Length 189791;
Best Local Similarity 92.0%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Db 115168 AAAAAAAAAATCCAGCAAAATCT 115192

Oy 1 AAAAAAAAAATCCAGCAAAATCT 25

RESULT 40

AC068681/c

LOCUS

DEFINITION

AC068681.2

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

ATTNORS

TITLE

JOURNAL

REFERENCE

ATTNORS

1 196831 bp DNA linear HTG 24-AUG-2002

Homo sapiens chromosome 6 clone RP11-671L8 map 6, WORKING DRAFT

SEQUENCE, 20 unordered pieces.

AC068681

HTG: HTGS PHASE1; HTGS_DRAFT.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 196831)

Birren,B., Linton,L., Nuebaum,C. and Lander,E.

Homo sapiens chromosome 6, clone RP11-671L8

Unpublished

2 (bases 1 to 196831)

Birren,B., Linton,L., Nuebaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,M., Burkett,G., Campoliano,A., Casale,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karstae,A., Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Margula,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., McDermid,J., Meneau,L., Mihov,T., Miranda,C., Mianga,V., Morrow,J., Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,R., Koltman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testafe,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J., Vassiliou,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wymann,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 196831)

Birren,B., Linton,L., Nuebaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,M., Burkett,G., Campoliano,A., Casale,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuirk,A., McKernan,C., McPheters,R.,
Meidrim,J., Menes,L., Mihova,T., Miranda,C., Mlenda,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,I., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Plesni,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talmas,J.,
Teetate,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vasilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2000 this sequence version replaced gi:7712158.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----
Center project name: LS642
Center clone name: 671.L.8

----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187164 bases at least Q40
Consensus quality: 191936 bases at least Q30
Consensus quality: 193826 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 194931; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1018: contig of 1018 bp in length
* 1019 1118: gap of 100 bp
* 1119 2873: contig of 1755 bp in length
* 2874 2973: gap of 100 bp
* 2974 4656: contig of 1683 bp in length
* 4657 4756: gap of 100 bp
* 4757 7018: contig of 2262 bp in length
* 7019 7118: gap of 100 bp
* 7119 9644: contig of 2526 bp in length
* 9645 9745: gap of 100 bp
* 9746 11714: contig of 1970 bp in length
* 11715 11814: gap of 100 bp
* 11815 14827: contig of 3013 bp in length
* 14828 14927: gap of 100 bp
* 14929 17471: contig of 2544 bp in length
* 17472 17571: gap of 100 bp
* 17572 21911: contig of 4330 bp in length
* 21912 22011: gap of 100 bp
* 22012 27134: contig of 5133 bp in length
* 27135 27234: gap of 100 bp
* 27235 32034: contig of 4800 bp in length
* 32035 32134: gap of 100 bp
* 32135 39278: contig of 7144 bp in length
* 39279 39378: gap of 100 bp

FEATURES
source

* 39379 44125: contig of 4747 bp in length
* 44126 44225: gap of 100 bp
* 44226 58909: contig of 14684 bp in length
* 58910 59009: gap of 100 bp
* 59010 75495: contig of 16486 bp in length
* 75496 75595: gap of 100 bp
* 75596 95743: contig of 20148 bp in length
* 95744 95844: gap of 100 bp
* 95844 119614: contig of 23771 bp in length
* 119615 119714: gap of 100 bp
* 119715 143111: contig of 23397 bp in length
* 143112 143211: gap of 100 bp
* 143212 171502: contig of 28291 bp in length
* 171503 171602: gap of 100 bp
* 171603 196831: contig of 25229 bp in length.

Location/Qualifiers
1..196831
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-671L8"
/clone_lib="RPCT-11 Human Male BAC"
1..1018
/note="assembly_fragment"
1019..1118
/estimated_length=100
1119..2873
/note="assembly_fragment"
2874..2973
/estimated_length=100
2974..4656
/note="assembly_fragment"
4657..4756
/estimated_length=100
4757..7018
/note="assembly_fragment"
7019..7118
/estimated_length=100
7119..9644
/note="assembly_fragment"
9645..9744
/estimated_length=100
9745..11714
/note="assembly_fragment"
11715..11814
/estimated_length=100
11815..14827
/note="assembly_fragment"
14828..14927
/estimated_length=100
14928..17471
/note="assembly_fragment"
17472..17571
/estimated_length=100
17572..21911
/note="assembly_fragment"
21912..22011
/estimated_length=100
22012..27134
/note="assembly_fragment"
27135..27234
/estimated_length=100
27235..32034
/note="assembly_fragment"
32035..32134
/estimated_length=100
32135..39278
/note="assembly_fragment"
39279..39378
/estimated_length=100
39379..44125

gap /note="assembly_fragment"
44126 . 44225
estimated_length=100
44226 . 58909
/note="assembly_fragment
clone_end=17

Query Match 87.2% Score 21.8; DB 14; Length 196831;
Best Local Similarity 92.0%; Pred No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAACTCT 25
Db 169813 AAAAAAAAAATCCAAACAACTCT 169789

RESULT 41
AC104343 203753 bp DNA linear PRI 23-MAR-2002
LOCUS Homo sapiens chromosome 8, clone RP11-822K17, complete sequence.
DEFINITION AC104343
AC104343.3 GI:19697489
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE 1 (bases 1 to 203753)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone RP11-822K17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 203753)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Baetsen, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
Choepl, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kelle, C., Lacroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McNetters, R., Meldrum, J.,
Menue, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Olivier, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Teffaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vasilev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Submitted (08-DEC-2001) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 203753)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Baetsen, V., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,
Chazaro, B., Choepl, Y., Colangelo, M., Collins, S., Collymore, A.,
Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kelle, C., Lacroque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menue, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,

TITLE Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Teffaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vasilev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

COMMENT Submitted (23-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2002 this sequence version replaced gi:18250045.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

FEATURES
source
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L20715
Center clone name: 822_K_17

Location/Qualifiers
1..203753
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-822K17"
/clone_id="RP11-822K17 Human Male BAC"
1..299
/rpt_family="THE1A"
531..610
/rpt_family="MIR"
complement(743..990)
/rpt_family="MIR"
2249..2521
/rpt_family="L1MBC"
2679..2819
/rpt_family="FLAM_C"
2862..2916
/rpt_family="L1ME3"
2928..3097
/rpt_family="L1ME3"
3098..3415
/rpt_family="A1UY"
3416..3610
/rpt_family="L1ME3"
3611..3652
/rpt_family="CA)n"
3656..3939
/rpt_family="L1ME3"
4346..4583
/rpt_family="AT_rich"
complement(5570..6301)
/rpt_family="ERVU"
6303..7267
/rpt_family="ERVU-C1"
7275..7654
/rpt_family="MLT2C1"
7755..7787
/rpt_family="AT_rich"
8871..9390
/rpt_family="MLT1G"
9459..9490
/rpt_family="AT_rich"
9664..10341
/rpt_family="THE1C"
10681..10701
/rpt_family="AT_rich"


```

repeat_region      10767..10848
                    /rpt_family="TA)n"
repeat_region      11419..11447
                    /rpt_family=" (CAA)n"
repeat_region      complement(12034..12183)
                    /rpt_family="MERS7A"
repeat_region      12305..12322
                    /rpt_family=" (TTG)n"
repeat_region      12470..12547
                    /rpt_family="L2"
repeat_region      13216..13470
                    /rpt_family="AluSg/x"
repeat_region      13916..14069
                    /rpt_family="MTR"
repeat_region      14339..14625
                    /rpt_family="L3"
repeat_region      complement(14990..15197)
                    /rpt_family="MTR"
repeat_region      complement(15198..15236)
                    /rpt_family="THE1C"
repeat_region      15237..15532
                    /rpt_family="AluSx"
repeat_region      complement(15533..15857)
                    /rpt_family="THE1C"
repeat_region      complement(15858..15900)
                    /rpt_family="MTR"
repeat_region      17545..17844
                    /rpt_family="AluSp"
repeat_region      17885..17932
                    /rpt_family="AT_rich"
repeat_region      complement(19265..19464)
                    /rpt_family="LMA9"
repeat_region      19465..19600
                    /rpt_family="LMA9"
repeat_region      19881..20254
                    /rpt_family="LMA9"
repeat_region      complement(20274..20628)
                    /rpt_family="LTPB1"
repeat_region      20699..26718
                    /rpt_family="LTPA2"
repeat_region      complement(26804..29602)
                    /rpt_family="Harlequin"
repeat_region      29603..29902
                    /rpt_family="AluY"
repeat_region      complement(29903..30280)
                    /rpt_family="Harlequin"
repeat_region      30282..30365
                    /rpt_family="LTR8"
repeat_region      complement(30368..31325)
                    /rpt_family="HERV1"
repeat_region      complement(31381..31468)
                    /rpt_family="HERV1"
repeat_region      31469..31545
                    /rpt_family="MERSB"
repeat_region      complement(31551..31774)
                    /rpt_family="HERV1"
repeat_region      complement(31776..33078)
                    /rpt_family="HERV1"
repeat_region      complement(33088..33550)
                    /rpt_family="LTR10A"
repeat_region      33562..34140
                    /rpt_family="LTPB1"
repeat_region      34149..39344
                    /rpt_family="LTPB1"
repeat_region      39345..39633
                    /rpt_family="AluY"
repeat_region      39634..39955
                    /rpt_family="LTPB1"
repeat_region      complement(40137..40372)
                    /rpt_family="MTR"
repeat_region      complement(40480..41043)
                    /rpt_family="L2"

```

```

Query Match      87.2%; Score 21.8; DB 8; Length 203753;
Best Local Similarity 92.0%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAATCGAGCAATCT 25
Db      58639 AAAAAAAAAATTGCGAATAATCT 58615

RESULT 42
AC165290
LOCUS
DEFINITION
AC165290
AC165290 BAC clone RP23-343A4 from chromosome 13, complete
sequence.
AC165290 AC103611
VERSION
AC165290.2 GI:71533453
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 206249)
lek.S. Kozlowicz,A. and Cotton,M.
The sequence of Mus musculus BAC clone RP23-343A4
Unpublished (2001)
2 (bases 1 to 206249)
Wilson,R.K.
Direct Submission
Submitted (09-JUL-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 206249)
Wilson,R.K.
Direct Submission
Submitted (30-JUL-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 30, 2005 this sequence version replaced gi:70668978.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submisions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_BA0343A04
Drafting center: WIBR

COMMENT

```

NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.

This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.

If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
The BAC library has been constructed by Kazutoyo Osegawa and
Mitsuko Tateo in the laboratory of Pieter de Jong
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (<http://www.reagen.com>) or
Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

FEATURES

Location/Qualifiers
1..206249
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="13"
/clone="RP23-343A4"
/clone_1b="RPCT-23"
15353..15481
/note="Sequence derived from PCR product of genomic DNA"
134725..135487
/note="Unresolved simple sequence repeat."

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 206249;
Best Local Similarity 92.0%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAAATCGAGCAAAATCT 25
31925 AAAAAAAAAATCCAGCCAAATCT 31949

Db 31925 AAAAAAAAAATCCAGCCAAATCT 31949

RESULT 43

LOCUS AC016311 224450 bp DNA linear HTG 16-APR-2000
DEFINITION Homo sapiens clone RP11-27H10, LOW-PASS SEQUENCE SAMPLING.
AC016311
AC016311.3 GI:7577571
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo

REFERENCE 1 (bases 1 to 224450)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-27H10
Unpublished
2 (bases 1 to 224450)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barne,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Caetle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,C., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Headford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kam,L., Karasas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., MacDonald,P., Margulis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.U., Zimmer,A. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 16, 2000 this sequence version replaced gi:6970658.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L4696
Center clone name: 27_H_10

NOTE: This record contains 246 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 727: contig of 727 bp in length
728 827: gap of 100 bp
828 1601: contig of 774 bp in length
1602 1701: gap of 100 bp
1702 2457: contig of 756 bp in length
2458 2557: gap of 100 bp
2558 3331: contig of 774 bp in length
3332 3431: gap of 100 bp
3432 4213: contig of 782 bp in length
4214 4313: gap of 100 bp
4314 5079: contig of 766 bp in length
5080 5179: gap of 100 bp
5180 5947: contig of 768 bp in length
5948 6047: gap of 100 bp
6048 6817: contig of 770 bp in length
6818 6917: gap of 100 bp
6918 7666: contig of 749 bp in length
7667 7766: gap of 100 bp
7767 8515: contig of 749 bp in length
8516 9385: contig of 770 bp in length
9386 9485: gap of 100 bp
9486 10258: contig of 773 bp in length
10259 10358: gap of 100 bp
10359 11138: contig of 780 bp in length
11139 11238: gap of 100 bp
11239 12014: contig of 776 bp in length
12015 12114: gap of 100 bp
12115 12876: contig of 762 bp in length
12877 12976: gap of 100 bp
12977 13757: contig of 781 bp in length
13758 13857: gap of 100 bp
13858 14618: contig of 761 bp in length
14619 14718: gap of 100 bp
14719 15475: contig of 757 bp in length
15476 15575: gap of 100 bp
15576 16341: contig of 766 bp in length
16342 16441: gap of 100 bp
16442 17185: contig of 744 bp in length
17186 17285: gap of 100 bp
17286 18056: contig of 771 bp in length
18057 18156: gap of 100 bp
18157 18933: contig of 776 bp in length
18934 19032: gap of 100 bp
19033 19798: contig of 766 bp in length
19799 19898: gap of 100 bp
19899 20670: contig of 772 bp in length
20671 20770: gap of 100 bp
20771 21545: contig of 775 bp in length
21546 21645: gap of 100 bp
21646 22400: contig of 755 bp in length
22401 22500: gap of 100 bp
22501 23268: contig of 768 bp in length
23269 23368: gap of 100 bp
23369 24138: contig of 770 bp in length

```

* 24139 24238: gap of 100 bp
* 24239 25003: contig of 765 bp in length
* 25004 25103: gap of 100 bp
* 25104 25865: contig of 762 bp in length
* 25866 25965: gap of 100 bp
* 25966 26725: contig of 760 bp in length
* 26726 26825: gap of 100 bp
* 26826 27599: contig of 774 bp in length
* 27600 27689: gap of 100 bp
* 27700 28500: contig of 801 bp in length
* 28501 29383: contig of 783 bp in length
* 29384 29483: gap of 100 bp
* 29484 30228: contig of 745 bp in length
* 30229 30328: gap of 100 bp
* 30329 31091: contig of 763 bp in length
* 31092 31191: gap of 100 bp
* 31192 31946: contig of 755 bp in length
* 31947 32046: gap of 100 bp
* 32047 32828: contig of 782 bp in length
* 32829 32928: gap of 100 bp
* 32929 33699: contig of 771 bp in length
* 33700 33799: gap of 100 bp
* 33800 34564: contig of 765 bp in length
* 34565 34664: gap of 100 bp
* 34665 35408: contig of 744 bp in length
* 35409 35508: gap of 100 bp
* 35509 36271: contig of 763 bp in length
* 36272 36371: gap of 100 bp
* 36372 37136: contig of 765 bp in length
* 37137 37236: gap of 100 bp
* 37237 38006: contig of 770 bp in length
* 38007 38106: gap of 100 bp
* 38107 38850: contig of 744 bp in length
* 38851 38950: gap of 100 bp
* 38951 39733: contig of 783 bp in length
* 39734 39833: gap of 100 bp
* 39834 40619: contig of 786 bp in length
* 40620 40719: gap of 100 bp
* 40720 41471: contig of 752 bp in length
* 41472 41571: gap of 100 bp
* 41572 42342: contig of 771 bp in length
* 42343 42442: gap of 100 bp
* 42443 43207: contig of 765 bp in length
* 43208 43307: gap of 100 bp
* 43308 44050: contig of 743 bp in length
* 44051 44150: gap of 100 bp
* 44151 44910: contig of 760 bp in length
* 44911 45010: gap of 100 bp
* 45011 45791: contig of 781 bp in length
* 45792 45891: gap of 100 bp
* 45892 46550: contig of 759 bp in length
* 46551 46750: gap of 100 bp
* 46751 47511: contig of 761 bp in length
* 47512 47611: gap of 100 bp
* 47612 48339: contig of 728 bp in length
* 48340 48439: gap of 100 bp
* 48440 49195: contig of 756 bp in length
* 49196 49295: gap of 100 bp
* 49296 50069: contig of 774 bp in length
* 50070 50169: gap of 100 bp
* 50170 50916: contig of 747 bp in length
* 50917 51016: gap of 100 bp
* 51017 51770: contig of 754 bp in length
* 51771 51870: gap of 100 bp
* 51871 52633: contig of 763 bp in length
* 52634 52733: gap of 100 bp
* 52734 53486: contig of 753 bp in length
* 53487 53586: gap of 100 bp
* 53587 54330: contig of 744 bp in length
* 54331 54430: gap of 100 bp
* 54431 55122: contig of 692 bp in length
* 55123 55222: gap of 100 bp

```

```

* 55223 55974: contig of 752 bp in length
* 55975 56074: gap of 100 bp
* 56075 56844: contig of 770 bp in length
* 56845 56944: gap of 100 bp
* 56945 57625: contig of 681 bp in length
* 57626 57725: gap of 100 bp
* 57726 58519: contig of 794 bp in length
* 58520 58619: gap of 100 bp
* 58620 59399: contig of 780 bp in length
* 59400 59400: gap of 100 bp
* 59500 60274: contig of 775 bp in length
* 60275 60374: gap of 100 bp
* 60375 61153: contig of 779 bp in length
* 61154 61253: gap of 100 bp
* 61254 62030: contig of 777 bp in length
* 62031 62130: gap of 100 bp

Query Match
Best Local Similarity 92.0%; Pred. NO. 2.4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAATCT 25
Db 64284 AAAAAAAAAATCGCAGCAATCT 64308

RESULT 44
AC016311/c
LOCUS
AC016311
DEFINITION
AC016311 Homo sapiens clone RP11-27H10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION
AC016311.3 GI:7577571
VERSION
HTG; HTGS PHASE0.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 224450)
REFERENCE
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
2 (bases 1 to 224450)
REFERENCE
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens, clone RP11-27H10
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 224450)
REFERENCE
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Casale, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kahn, B., Karatas, A., Klein, J.,
Leloczky, J., Lien, C., Locke, K., Macdonald, P., Margulis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirelli, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 16, 2000 this sequence version replaced gi:6970658.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L4696
Center clone name: 27_H_10

```

* NOTE: This record contains 246 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
728 827: contig of 727 bp in length
828 827: gap of 100 bp
1602 1601: contig of 774 bp in length
1702 1701: gap of 100 bp
2458 2457: contig of 756 bp in length
2558 2557: gap of 100 bp
3332 3331: contig of 774 bp in length
3432 3431: gap of 100 bp
4214 4213: contig of 782 bp in length
4314 4313: gap of 100 bp
5080 5079: contig of 766 bp in length
5180 5179: gap of 100 bp
5948 5947: contig of 768 bp in length
6048 6047: gap of 100 bp
6818 6817: contig of 770 bp in length
6918 6917: gap of 100 bp
7667 7666: contig of 749 bp in length
7767 7766: gap of 100 bp
8516 8515: contig of 749 bp in length
8616 8615: gap of 100 bp
9386 9385: contig of 770 bp in length
9486 9485: gap of 100 bp
10259 10258: contig of 773 bp in length
10359 10358: gap of 100 bp
11139 11138: contig of 780 bp in length
11239 11238: gap of 100 bp
12015 12014: contig of 776 bp in length
12115 12114: gap of 100 bp
12876 12875: contig of 762 bp in length
12877 12876: gap of 100 bp
13758 13757: contig of 781 bp in length
13858 13857: gap of 100 bp
14619 14618: contig of 761 bp in length
14719 14718: gap of 100 bp
15476 15475: contig of 757 bp in length
15576 15575: gap of 100 bp
16342 16341: contig of 766 bp in length
16442 16441: gap of 100 bp
17186 17185: contig of 744 bp in length
17286 17285: gap of 100 bp
18057 18056: contig of 771 bp in length
18157 18156: gap of 100 bp
18933 18932: contig of 776 bp in length
19033 19032: gap of 100 bp
19799 19798: contig of 766 bp in length
19899 19898: gap of 100 bp
20671 20670: contig of 772 bp in length
20771 20770: gap of 100 bp
21545 21544: contig of 775 bp in length
21646 21645: gap of 100 bp
22401 22400: contig of 755 bp in length
22501 22500: gap of 100 bp
23269 23268: contig of 768 bp in length
23369 23368: gap of 100 bp
24139 24138: contig of 770 bp in length
24239 24238: gap of 100 bp
25004 25003: contig of 765 bp in length
25104 25103: gap of 100 bp
25866 25865: contig of 762 bp in length
25966 25965: gap of 100 bp
26726 26725: contig of 760 bp in length
26826 26825: gap of 100 bp

26826 27599: contig of 774 bp in length
27600 27599: gap of 100 bp
27700 28500: contig of 801 bp in length
28501 28600: gap of 100 bp
28601 29383: contig of 783 bp in length
29384 29483: gap of 100 bp
29484 30228: contig of 745 bp in length
30229 30328: gap of 100 bp
30329 31091: contig of 763 bp in length
31092 31191: gap of 100 bp
31192 31946: contig of 755 bp in length
31947 32046: gap of 100 bp
32047 32828: contig of 782 bp in length
32829 32928: gap of 100 bp
32929 33699: contig of 771 bp in length
33700 33799: gap of 100 bp
33800 34564: contig of 765 bp in length
34565 34664: gap of 100 bp
34665 35408: contig of 744 bp in length
35409 35508: gap of 100 bp
35509 36271: contig of 763 bp in length
36272 36371: gap of 100 bp
37137 37136: contig of 765 bp in length
37237 37236: gap of 100 bp
38007 38006: contig of 770 bp in length
38107 38106: gap of 100 bp
38851 38850: contig of 744 bp in length
38951 39733: contig of 783 bp in length
39734 39833: gap of 100 bp
39834 40619: contig of 786 bp in length
40620 40719: gap of 100 bp
40720 41471: contig of 752 bp in length
41472 41571: gap of 100 bp
41572 42342: contig of 771 bp in length
42343 42442: gap of 100 bp
42443 43207: contig of 765 bp in length
43208 43307: gap of 100 bp
43308 44050: contig of 743 bp in length
44051 44150: gap of 100 bp
44151 44910: contig of 760 bp in length
44911 45010: gap of 100 bp
45011 45791: contig of 781 bp in length
45792 45891: gap of 100 bp
45892 46690: contig of 759 bp in length
46691 46750: gap of 100 bp
46751 47511: contig of 761 bp in length
47512 47611: gap of 100 bp
47613 48339: contig of 728 bp in length
48340 48439: gap of 100 bp
48440 49195: contig of 756 bp in length
49196 49295: gap of 100 bp
49296 50069: contig of 774 bp in length
50070 50169: gap of 100 bp
50170 50916: contig of 747 bp in length
50917 51016: gap of 100 bp
51017 51770: contig of 754 bp in length
51771 51870: gap of 100 bp
51871 52633: contig of 763 bp in length
52634 52733: gap of 100 bp
52734 53486: contig of 753 bp in length
53487 53586: gap of 100 bp
53587 54330: contig of 744 bp in length
54331 54430: gap of 100 bp
54431 55122: contig of 692 bp in length
55123 55222: gap of 100 bp
55223 55974: contig of 752 bp in length
55975 56074: gap of 100 bp
56075 56844: contig of 770 bp in length
56845 56944: gap of 100 bp
56945 57625: contig of 681 bp in length
57626 58519: gap of 100 bp
58519 58519: contig of 794 bp in length

```

* 58520 58619: gap of 100 bp
* 58620 59399: contig of 780 bp in length
* 59400 59499: gap of 100 bp
* 59500 60274: contig of 775 bp in length
* 60375 60375: gap of 100 bp
* 61153 61153: contig of 779 bp in length
* 61254 61253: gap of 100 bp
* 61254 62030: contig of 777 bp in length
* 62031 62130: gap of 100 bp

Query Match      87.2% Score 21.8; DB 14; Length 224450;
Best Local Similarity 92.0%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACGACCAATCT 25
Db 194741 AAAAAAAAAATTCGACCAATCT 194717

RESULT 45
AC115438 231195 bp DNA linear HTG 08-OCT-2002
LOCUS Rattus norvegicus clone CH230-18982, *** SEQUENCING IN PROGRESS
DEFINITION *** 2 unordered pieces.
AC115438.4 GI:22856899
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 231195)
Muzny D.Marie, Metzker M.Lee., Abramson S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Angilano D.,
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Biewald K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Cessar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., P'Souza L.,
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Gebregorjais B., Geer K., Gill R., Grady M., Guerra W., Guevara W.,
Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hernandez R., Hines S., Hladun S.L., Hodgson A., Hognes M.,
Hollins B., Howells S., Huylk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Kacpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
Kowals C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorenshewer L., Loulseged H., Lozano R.J., Lu X., Ma J.,
Maheshwari M., Mahindaratne M., Mahmoud M., Malloy K., Mangum A.,
Mangum B., Mapua P., Martin K., Martin R., Martinez E.,
Mawliny S., McLeod M.P., McNeill T.Z., Meenen B.,
Milosavljevic A., Miner G., Mnja E., Montemayor J., Moore S.,
Morjan M., Morris K., Morris S., Munidasa M., Murphy M., Natir L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norrie S.,
Nwackemehe O., Okunonu G., Olarnunpong A., Pal S., Parks K.,
Pasternak S., Paul H., Perez A., Perez L., Pfankuch C.,
Plopper F., Poindecker A., Popovic D., Primus E., Pu L., Lu.,
Puafo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
Sanders W., Savery G., Scherer S., Scott G., Shatman S., Shen H.,
Shetty J., Shvartsbeyn A., Sison I., Sitter C.D., Smajls D.,
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,

```

```

TITLE JOURNAL
REFERENCE JOURNAL
AUTHORS JOURNAL
JOURNAL
REFERENCE JOURNAL
AUTHORS JOURNAL
JOURNAL
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GPWE
Center clone name: CH230-18982
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 21416 bases at least Q40
Consensus quality: 217873 bases at least Q20
Estimated insert size: 232141; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 231923: contig of 231923 bp in length
231924 232023: gap of unknown length
232024 233195: contig of 1172 bp in length.
Location/Qualifiers
1. 233195
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-18982"
231924..232023
/estimated_length=unknown
FEATURES
source
gap

```

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 233195;
 Best Local Similarity 92.0%; Pred. No. 2.4e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAAATCT 25
 |||||
 Db 91047 AAAAAAAAAATCGACAAATCT 91071

RESULT 46
 BS000230 242704 bp DNA linear PRI 12-JUN-2004
 LOCUS Pan troglodytes chromosome 22 clone:PTB-105H12, map 22, complete
 DEFINITION
 ACCESSION BS000230 BA000046
 VERSION BS000230.1 GI:37537497
 KEYWORDS HTG.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Pan.

REFERENCE
 1 The International Chimpanzee Chromosome 22 Consortium.
 DNA sequence and comparative analysis of chimpanzee chromosome 22
 Nature 429, 382-388 (2004)
 2 (bases 1 to 242704)
 Tsai, S., Liu, T., Wu, K., Liao, T. and Hsiao, K.
 Direct Submission
 Submitted (16-MAY-2003) Shih-Feng Tsai, National Health Research
 Institutes (NHRI), Division of Molecular and Genetic Medicine, 128,
 Yen-Chiu-Yuan Road, Sec 2, Taipei 115, Taiwan
 (E-mail: petesai@nhri.org.tw, URL: http://www.nhri.org.tw/
 Tel: 886-2-28267319, Fax: 886-2-28200552)
 The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
 *Chinese National Human Genome Center at Shanghai, Shanghai, China;
 *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
 of Molecular Biotechnology, Jena, Germany; *KIBB Genome Research
 Center, Daejeon, Korea;
 *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
 *National Institute of Genetics, Mishima, Japan;
 *National Yang Ming University Genome Research Center, Taipei,
 Taiwan;
 *RIKEN Genomic Sciences Center, Yokohama, Japan.

----- Genome Center
 Center: National Yang Ming University Genome Research Center
 code: YMG
 Web site: http://genome.ym.edu.tw/
 Contact: sequencem@ym.edu.tw

----- Project Information
 Center project name: The Chimpanzee Chromosome 22 Sequencing Project
 Center clone name: HC

----- Summary Statistics
 Sequencing vector: pUC18; 100% of reads
 Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly
 program: Phrap; version 0.990319
 Consensus quality: 242,677 bases at least Q40
 Consensus quality: 242,704 bases at least Q30
 Consensus quality: 242,704 bases at least Q20

 This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by at one
 plasmid
 subclone or more than one M13 subclone;
 and the assembly was confirmed by restriction digest.

 Source information:

The PTB1 chimpanzee BAC library was prepared from DNA isolated from
 cultured cells established from the blood of a single male
 chimpanzee.
 Clones may be obtained from Asao Fujiyama and co-workers
 (http://www.gsc.riken.go.jp).
 VECTOR: pKS145

 Sequence Quality Assessment:
 This entry has been annotated with sequence
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in
 10,000 bp.

 Neighboring clones: RP43-015P20 (left) and PTB-071B01 (right).
 Location/Qualifiers
 1. 242704
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="22"
 /clone="PTB-105H12"
 /clone_1kb="PTB1 chimpanzee BAC"

FEATURES
 source

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 242704;
 Best Local Similarity 92.0%; Pred. No. 2.4e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAAATCT 25
 |||||
 Db 145859 AAAAAAAAAATCGACAAATCT 145883

RESULT 47
 AC128115
 LOCUS Rattus norvegicus clone CH230-411K1, WORKING DRAFT SEQUENCE.
 DEFINITION
 ACCESSION AC128115.3 GI:25007802
 VERSION
 KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.

REFERENCE
 1 Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alshrooke, S., Amin, A., Angiano, D.,
 Ayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Bialwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Canter, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy, Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Faller, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geier, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howell, S., Hu, Y., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karachny, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovat, C.,
 Karpach, S., Kraft, C.L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshewa, L., Louieged, H., Lozado, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,

```

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundaas, M., Murphy, M., Nair, L.,
Narkervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokoemeh, O., Okunnu, G., Olarnunagoon, A., Pal, S., Parke, K.,
Paserniak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plapper, F., Poindexter, A., Popovic, D., Pims, E., Pu, L.,
Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shneyd, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smejs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Treloar, Z., Ueman, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 280569)
Worley, K.C.
Direct Submission
Submitted (19-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 280569)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23908294.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZPX
Center clone name: CH230-411K1
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 145762 bases at least Q40
Consensus quality: 148026 bases at least Q30
Consensus quality: 148813 bases at least Q20
Estimated insert size: 148794; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have

```

```

* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 280569: contig of 280569 bp in length.
Location/Qualifiers
1..280569
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-411K1"

```

```

misc_feature
1..1237
/note="wgs_end_extension
clone_end:Sp6"
105410..106766
/note="wgs_end_extension
clone_end:Sp6"
107611..109266
/note="wgs_end_extension
clone_end:Sp6"
complement(109556..110506)
/note="clone boundary
clone_end:Sp6
site:
end_sequence:B2118183"
273438..274298
/note="clone boundary
clone_end:T7
site:
end_sequence:B2117765"
277189..280569
/note="wgs_end_extension
clone_end:T7"

```

```

ORIGIN
Query Match 87.2%; Score 21.8; DB 14; Length 280569;
Best Local Similarity 92.0%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Db 1 AAAAAAAAAATCGACAAATCT 25
260526 AAAAAAAAAACTTCGACAAATCT 260550

```

```

RESULT 48
AC024147
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-199N2, WORKING DRAFT SEQUENCE,
33 unordered pieces.
ACCESSION
AC024147
VERSION
AC024147.16 GI:21426080
KEYWORDS
HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 299744)

```

```

REFERENCE
1 (bases 1 to 299744)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C.,
Alabrooks, S.L., Amaralunge, H.C., Are, J.R., Ayala, M., Banks, T.,
Barbarta, J., Benton, J., Bimage, K., Blankenburg, K., Bonin, D.,
Bouck, J., Boyle, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Dayila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinb, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Geo, U., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

```

```

Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
Hornel, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Joliver, S., Joudah, S.,
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kretovc, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegod, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Magua, P., Martin, R., Marindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moore, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokkenwo, S., Ogun, M., Okunodu, G.,
Oragunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Petere, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scheier, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Syatek, A., Taber, P., Tamerias, A., Tamerias, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 299744)
Worley, K.C.
Direct Submission
Submitted (25-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 299744)

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: HAGW
Center clone name: RP11-199N2
-----
Summary Statistics
Sequencing vector: M13
Chemistry: Dye-Primer Body: 8% of reads
Assembly: Dye-terminator Big Dye: 92% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 273645 bases at least Q40
Consensus quality: 282517 bases at least Q30
Consensus quality: 287509 bases at least Q20
Remapped insert size: 295437; sum-of-coverage estimation
Quality coverage: 4x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a "working draft" sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2079: contig of 2079 bp in length
* 2080 2179: gap of unknown length
* 2080 2179: gap of unknown length

```

```

* 2180 4411: contig of 2232 bp in length
* 4412 4511: gap of unknown length
* 4512 6727: contig of 2216 bp in length
* 6728 6827: gap of unknown length
* 6828 9805: contig of 2978 bp in length
* 9806 9905: gap of unknown length
* 9906 12355: contig of 2450 bp in length
* 12356 12455: gap of unknown length
* 12456 16839: contig of 4384 bp in length
* 16840 16939: gap of unknown length
* 16940 21020: contig of 4081 bp in length
* 21021 21120: gap of unknown length
* 21121 27628: contig of 6508 bp in length
* 27629 27728: gap of unknown length
* 27729 32949: contig of 5221 bp in length
* 32950 33049: gap of unknown length
* 33050 39837: contig of 6788 bp in length
* 39838 39937: gap of unknown length
* 39938 45918: contig of 5981 bp in length
* 45919 46018: gap of unknown length
* 46019 50103: contig of 4085 bp in length
* 50104 50203: gap of unknown length
* 50204 54390: contig of 4187 bp in length
* 54391 54490: gap of unknown length
* 54491 60088: contig of 5598 bp in length
* 60089 60188: gap of unknown length
* 60189 65783: contig of 5595 bp in length
* 65784 65883: gap of unknown length
* 65884 71998: contig of 6115 bp in length
* 71999 72098: gap of unknown length
* 72099 77751: contig of 5653 bp in length
* 77752 77851: gap of unknown length
* 77852 83677: contig of 5826 bp in length
* 83678 83777: gap of unknown length
* 83778 91814: contig of 8037 bp in length
* 91815 91914: gap of unknown length
* 91915 103030: contig of 11116 bp in length
* 103031 103130: gap of unknown length
* 103131 112041: contig of 8911 bp in length
* 112042 121041: gap of unknown length
* 121043 123072: contig of 10931 bp in length
* 123073 123172: gap of unknown length
* 123173 135138: contig of 11966 bp in length
* 135139 135238: gap of unknown length
* 135239 146265: contig of 11027 bp in length
* 146266 146365: gap of unknown length
* 146366 159132: contig of 12767 bp in length
* 159133 159233: gap of unknown length
* 159234 159369: contig of 10137 bp in length
* 159370 169469: gap of unknown length
* 169470 182352: contig of 12883 bp in length
* 182353 182452: gap of unknown length
* 182453 195298: contig of 12846 bp in length
* 195299 195398: gap of unknown length
* 195399 211343: contig of 15945 bp in length
* 211344 211443: gap of unknown length
* 211444 225262: contig of 13819 bp in length
* 225263 225362: gap of unknown length
* 225363 240775: contig of 15433 bp in length
* 240776 240875: gap of unknown length
* 240876 266463: contig of 25588 bp in length
* 266464 266563: gap of unknown length
* 266564 299744: contig of 33181 bp in length.
Location/Qualifile
1. .299744
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-199N2"
2080. .2179
/estimated_length=unknown
4412. .4511

```

```

FEATURES
source
gap
gap
gap

```


gap	/estimated_length=unknown 6728. .6827	
gap	/estimated_length=unknown 9806. .9905	
gap	/estimated_length=unknown 12356. .12455	
gap	/estimated_length=unknown 16840. .16939	
gap	/estimated_length=unknown 21021. .21120	
gap	/estimated_length=unknown 27629. .27728	
gap	/estimated_length=unknown 32950. .33049	
gap	/estimated_length=unknown 39838. .39937	
gap	/estimated_length=unknown 45919. .46018	
gap	/estimated_length=unknown 50104. .50203	
gap	/estimated_length=unknown 54391. .54490	
gap	/estimated_length=unknown 60089. .60188	
gap	/estimated_length=unknown 65784. .65883	
gap	/estimated_length=unknown	
Best Match	87.2%; Score 21.8; DB 14;	Length 299744;
Query Local Similarity	92.0%;	Pred. No. 2.3e+02;
Matches 23; Conservative 0;	Mismatches 2;	Indels 0; Gaps 0

```
QY      1 AAAAAAAAAATCGCAGCAATCT 25
          |||||
Db    136941 AAAAAAAAAACGCAGCAATAT 136965
```

RESULT 49	AC110996/c	LOCUS	DEFINITION	AC110996	306130 bp	DNA	linear	HTG 09-MAY-2007
AC110996/c			Homio sapiens chromosome 3 clone RP11-267D9, *** SEQUENCING IN PROGRESS ***; 46 unordered pieces.	AC110996				
AC110996			HTG: HTGS PHASE1.	AC110996.2	GI:20335984			
ORGANISM			Homio sapiens (human)					
REFERENCE			Homio sapiens					
AUTHORS			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.					
			1 (bases 1 to 306130)					
			Muzny D.M., Adams C., Adio-Odola, B., Ali-osman, F.R., Allen, C., Alibrooks, S.L., Amaralunge, H.C., Are, J.R., Ayale, W., Banks, T., Barbarta, J., Benton, J., Blinage, K., Blankenburg, K., Bonnin, D., Bouch, J., Bowle, S., Brieva, M., Brown, M., Brown, M., Bryant, N.P., Bunyak, C., Buruch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Deedrich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Dublin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, V.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, A., Hughes, J., Huijck, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudh, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Koyar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L., Lewis, L., Li, J., Li, Z., Lichtearge, O., Lileu, C., Liu, J., Liu, W., Louisse, H.,					

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 306130)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (17-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 306130)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Apr 28, 2002 this sequence version replaced gi:18698792.

Lozada,R.J.,Lu,X.,Lucier,A., Lucier,R., Luna,R., Ma,J.,
Mashwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Medendorp,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwendo,S., Ogih,M., Okunou,G.,
Oragunye,N., Orieden,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivera,M., Rojars,A., Rojboxkan,I., Rolfe,M., Ruiz,S., Savery,G.,
Schetter,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabors,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Teifrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleciyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 306130)
Worley,K.C.

Direct Submission
Submitted (17-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 306130)
Worley,K.C.

Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:18698792.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HDW
Center clone name: RP11-267D9

Summary Statistics
Sequencing vector: plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 345014 bases at least Q40
Consensus quality: 365949 bases at least Q30
Consensus quality: 382900 bases at least Q20
Estimated insert size: 18643; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
NOTE: This is a 'working draft' sequence. It currently
consists of 46 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
2101: contig of 2101 bp in length
2102
2201: gap of unknown length
2202
4326: contig of 2025 bp in length
4327
4327: gap of unknown length
4327
6365: contig of 2039 bp in length
6366
6465: gap of unknown length
6466
9390: contig of 2925 bp in length
9391
9490: gap of unknown length
9491
11753: contig of 2263 bp in length
11754
11853: gap of unknown length
11854
14201: contig of 2348 bp in length
14202
14301: gap of unknown length

14302	15506:	contig of 2205 bp in length
14607	15606:	gap of unknown length
15607	15606:	contig of 2300 bp in length
18907	15006:	gap of unknown length
19007	21027:	contig of 2021 bp in length
21028	21127:	gap of unknown length
21128	22324:	contig of 2107 bp in length
22325	22334:	gap of unknown length
22335	22853:	contig of 2519 bp in length
22854	22953:	gap of unknown length
22954	22913:	contig of 3660 bp in length
229614	229713:	gap of unknown length
229714	32739:	contig of 3026 bp in length
32740	32839:	gap of unknown length
32840	35845:	contig of 3006 bp in length
35846	35945:	gap of unknown length
35946	39274:	contig of 3329 bp in length
39275	39374:	gap of unknown length
39375	43568:	contig of 4194 bp in length
43569	43668:	gap of unknown length
43669	44554:	contig of 2766 bp in length
44555	44554:	gap of unknown length
44555	50988:	contig of 4434 bp in length
50989	51088:	gap of unknown length
51089	54919:	contig of 3831 bp in length
54920	55019:	gap of unknown length
55020	57982:	contig of 2963 bp in length
57993	58082:	gap of unknown length
58093	61464:	contig of 3382 bp in length
61465	61564:	gap of unknown length
61565	65184:	contig of 4620 bp in length
65185	66284:	gap of unknown length
66285	70187:	contig of 3903 bp in length
70188	70287:	gap of unknown length
70288	72468:	contig of 3181 bp in length
72469	73568:	gap of unknown length
73569	79809:	contig of 6241 bp in length
79810	79909:	gap of unknown length
85246	85263:	contig of 5334 bp in length
85246	85343:	gap of unknown length
85344	88547:	contig of 3184 bp in length
88548	88647:	gap of unknown length
88648	93863:	contig of 5216 bp in length
93864	93963:	gap of unknown length
93964	97505:	contig of 3542 bp in length
97506	97605:	gap of unknown length
97606	100406:	contig of 2801 bp in length
100407	100506:	gap of unknown length
100507	105444:	contig of 4938 bp in length
105445	105544:	gap of unknown length
105545	111235:	contig of 5691 bp in length
111236	111335:	gap of unknown length
111336	120553:	contig of 9218 bp in length
120554	120653:	gap of unknown length
120654	128040:	contig of 7387 bp in length
128041	128140:	gap of unknown length
128141	137122:	contig of 8892 bp in length
137123	137222:	gap of unknown length
137223	142844:	contig of 5626 bp in length
142845	142944:	gap of unknown length
142945	150181:	contig of 7233 bp in length
150182	150281:	gap of unknown length
150282	157958:	contig of 7677 bp in length
157959	158058:	gap of unknown length
158059	171797:	gap of unknown length
171798	171897:	gap of unknown length
171898	184421:	contig of 13524 bp in length
184422	184521:	gap of unknown length
184522	203377:	contig of 18852 bp in length
203377	219622:	gap of unknown length
203474	219622:	contig of 16149 bp in length
219623	219722:	gap of unknown length
219723	228092:	contig of 18371 bp in length

	* 238094	238193:	gap of unknown length	
	* 238194	257985:	contig of 19692 bp in length	
	* 257986	257985:	gap of unknown length	
	* 257986	281157:	contig of 23172 bp in length	
	*	281158	281257: gap of unknown length	
	*	281258	306130: contig of 24873 bp in length.	
FEATURES				
source				
		1..306130		
		/organism="Homo sapiens"		
		/mol_type="genomic DNA"		
		/db_xref="taxon:9606"		
		/chromosome="3"		
		/clone="RP11-267D9"		
gap		2102..2201		
gap		/estimated_length=unknown		
gap		4227..4326		
gap		/estimated_length=unknown		
gap		6366..6465		
gap		/estimated_length=unknown		
gap		9391..9490		
Query Match		87.2%:	Score 21.8;	DB 14; Length 306130;
Beet Local Similarity		92.0%:	Pred. No. 2.3e+02;	
Matches	23;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
Dn	205420	AAAAAAAAAATCCAGACAATCT 25		
Qy	1	AAAAAAAAAATCCAGACAATCT 25		
RESULT 50				
AC107171/c				
LOCUS				
DEFINITION	Rattus norvegicus clone cM230-203L1,	*** SEQUENCING IN PROGRESS		
ACCESSION	AC107171.4	GI:32321678		
VERSION	AC107171.4	GI:32321678		
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;			
	Sciurognathi; Muridae; Murinae; Rattus.			
	1 (bases 1 to 325493)			
REFERENCE	Muzny,D.,Marler,M.,Metzker,M.,Lee,S.,Abramson,S.,Adams,C.,Alder,J.,			
AUTHORS	Allen,C.,Allen,H.,Albrooke,S.,Amiri,A.,Angiano,D.,			
	Aryal-Becchi,V.,Ayodeji,A.,Ayodeji,I.W.,Baca,E.,Baden,H.,			
	Baldwin,D.,Bandarinalake,D.,Barber,M.,Barnstead,M.,Benahmed,F.,			
	Biswalto,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,			
	Bryan,N.,Buhaq,C.,Burck,C.,Buttelli,K.,Calderson,E.,			
	Cardenas,V.,Cartier,K.,Cavazos,I.,Ceasar,H.,Center,A.,			
	Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,			
	Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,			
	Davila,M.L.,Davis,C.,Davy-Carrillo,L.,De Anda,C.,Dederich,D.,			
	Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K.,			
	Diaper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,			
	Egan,A.,Escoto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,			
	Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foister,M.,Foister,P.,			
	Frazer,C.M.,Gabriel,A.,Ganta,R.,Garca,A.,Garnier,T.,Gaza,M.,			
	Gebremegiste,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,			
	Gumartne,P.,Haaland,W.,Hambl,C.,Hamilton,C.,Hamilton,K.,			
	Harvey,Y.,Havlik,P.,Hawes,A.,Henderson,N.,Hernandez,J.,			
	Hernandez,R.,Hines,S.,Hiadun,S.L.,Hodgeson,A.,Hogues,M.,			
	Hollins,B.,Howells,S.,Huily,S.,Hume,J.,Idlebird,D.,Jackson,A.,			
	Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jolyvet,A.,			
	Kapachay,S.,Kelly,S.,Khan,Z.,Kling,L.,Kovar,C.,			
	Kwais,C.,Kraft,C.L.,Lebow,H.,Levan,J.,Lewis,L.,Li,Z.,Liu,J.,			
	Liu,J.,Liu,W.,Liu,Y.,London,P.,Longacre,S.,Lopez,J.,			
	Lorenzenhuwa,L.,Louisleged,H.,Locado,R.J.,Lu,X.,Ma,J.,			
	Mathewswati,M.,Mahindaratne,M.,Mahmoud,M.,Malloy,C.,Manungu,A.,			
	Manungu,B.,Mapua,P.,Martin,K.,Martin,M.,Martinez,E.,			
	Mathiney,S.,McLeod,M.P.,McNeill,T.Z.,Meenen,E.,			
	Milosavljevic,A.,Miner,G.,Mitla,E.,Montemavor,J.,Moore,S.,			

```

Morgan,M., Morris,K., Morris,S., Munidaa,M., Murphy,M., Nair,L.,
Nanervlis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwackemele,O., Okwunou,G., Olarnunasegon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Ploner,F., Polindexter,A., Popovic,D., Primus,E., Pu,J.,-L.,
Puaa,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Rutz,S.D.,
Sander,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soza,J.,
Steinle,M., Strong,R., Sutton,S., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umant,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 325493)
Worley,K.C.
Direct Submission
Submitted (16-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 325493)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 26, 2002 this sequence version replaced gi:21744091.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using AClas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.

-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----Project Information
Center project name: GPPN
Center clone name: CH230-203L1
-----Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 251669 bases at least Q40
Consensus quality: 256472 bases at least Q40
Consensus quality: 260077 bases at least Q20
Estimated insert size: 284614; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 10786: contig of 10786 bp in length
* 10787 10886: gap of unknown length
* 10887 116797: contig of 105911 bp in length
* 116798 116897: gap of unknown length

FEATURES
source
1..325493
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-203L1"
1..1283
/note="wgs_end_extension
clone_end:Sp6
/clone="clone boundary
clone_end:Sp6
site:ECORI
end_sequence:BH364111"
9829..10339
/note="clone boundary
clone_end:T7
site:ECORI
end_sequence:BH364110"
10787..10886
/estimated_length=unknown
10887..112846
/note="wgs_end_extension
clone_end:T7"
15098..16436
/note="wgs_end_extension
clone_end:T7"
116798..116897
/estimated_length=unknown
189672..189771
/estimated_length=unknown
189772..190814
/note="wgs_end_extension
clone_end:T7"
196394..197511
/note="wgs_end_extension
clone_end:T7"
225556..225655
/estimated_length=unknown
242857..242956
/estimated_length=unknown
242957..245016
/note="wgs_end_extension
clone_end:T7"
256757..256856
/estimated_length=unknown
256857..258093
/note="wgs_end_extension
clone_end:T7"
275640..275739

```

```

gap      /estimated_length=unknown
          313980. .314079
          /estimated_length=unknown
gap      316290. .316389
          /estimated_length=unknown
gap      317993. .318092
          /estimated_length=unknown
gap      320005. .320104
          /estimated_length=unknown
gap      321585. .321684
          /estimated_length=unknown
gap      323477. .323576
          /estimated_length=unknown

```

```

Query Match      87.2%; Score 21.8; DB 14; Length 325493;
Best Local Similarity 92.0%; Pred. No. 2.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy      1 AAAAAAAAAATCGCAGACAAATCT 25
          |||||
Db      140408 AAAAAAAAACTCTCAGACAAATCT 140384

```

Search completed: December 14, 2005, 11:10:15
 Job time : 874.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 ; Search time 203.2 Seconds

(without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-2

Perfect score: 25

Sequence: 1 aaaaaaaaaatcgagacaatct 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 150 summaries

Database :

N_Geneseq_21:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002as:*

7: geneseqn2002bs:*

8: geneseqn2003as:*

9: geneseqn2003bs:*

10: geneseqn2003cs:*

11: geneseqn2003ds:*

12: geneseqn2004as:*

13: geneseqn2004bs:*

14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.8	87.2	142318	11 ACN44850	Acn44850 Human gen
2	20.4	81.6	4068	4 AAI86707	Aai86707 Human pol
3	20.4	81.6	7300	14 AD280551	Ad280551 SH3-domai
4	20.4	81.6	13388	10 ADC71368	Adc71368 Human col
5	20.4	81.6	22970	10 ADK70082	Adk70082 Murcan hu
6	20.4	81.6	22970	10 ADK70081	Adk70081 Wild type
7	20.4	81.6	241748	14 AD213116	Ad213116 Murine ca
8	20.2	80.8	1532	3 AAC49529	Aac49529 Arabidops
9	20.2	80.8	1536	3 AAC42454	Aac42454 Arabidops
10	20.2	80.8	3875	4 ABL28990	Abi28990 Drosophil
11	20.2	80.8	6162	6 ABR43074	Abi43074 Genomic s
12	20.2	80.8	6226	9 ADB61230	Adb61230 Connectiv
13	20.2	80.8	6226	9 ADB61230	Adb61230 Connectiv
14	20.2	80.8	7145	6 ABU92234	Abu92234 ChemicalI
15	20.2	80.8	7145	6 AAD22321	Aad22321 ChemicalI
16	20.2	80.8	7928	6 ABL32095	Abi32095 Human imm
17	20.2	80.8	7928	6 ABK31169	Abk31169 Signal tr
18	20.2	80.8	7928	6 ABL70124	Abi70124 ChemicalI
19	20.2	80.8	7928	6 AAB61056	Aab61056 Human gen

20	20.2	80.8	17677	4 AAK84095	Aak84095 Human imm
21	20.2	80.8	17677	4 AAK70506	Aak70506 Human imm
22	20.2	80.8	29105	9 AAD56081	Aad56081 Human MAP
23	20.2	80.8	29105	9 ADA02443	Ada02443 Human MAP
24	20.2	80.8	29105	10 ADB72182	Adb72182 Human MAP
25	20.2	80.8	31024	10 ABV75372	Abv75372 Human IGF
26	20.2	80.8	37736	4 AAK86139	Aak86139 Human imm
27	20.2	80.8	41936	6 ABL67924	Abi67924 Ovary can
28	20.2	80.8	110000	12 ADJ25985_11	Continuation (12 o
29	20.2	80.8	110000	12 ADN97989_11	Continuation (12 o
30	20.2	80.8	110000	12 ADOS0281_11	Continuation (12 o
31	20.2	80.8	110000	14 AEB85185_11	Continuation (12 o
32	20.2	80.8	110000	14 AEB85185_11	Continuation (12 o
33	20.2	80.8	152759	13 ABD33075	Abd33075 Human can
34	20.2	80.8	180227	13 ABD33268	Abd33268 Human can
35	19.8	79.2	230	7 ADS31291	AdS31291 Human gen
36	19.8	79.2	230	7 ADY36679	Ady36679 HIRA geno
37	19.8	79.2	267	13 ACN51137	Acn51137 Corton an
38	19.8	79.2	577	6 AEO21911	Abq21911 OligonucI
39	19.8	79.2	577	6 AEO21910	Abq21910 OligonucI
40	19.8	79.2	1710	6 ABL34385	Abi34385 Human imm
41	19.8	79.2	5596	2 AAV83941	Aav83941 Bacterial
42	19.8	79.2	40783	12 ADO97173	Adq97173 Human can
43	19.8	79.2	80240	2 AAV83940	Aav83940 NC-contig
44	19.8	79.2	80595	2 AAV83939	Aav83939 HC-contig
45	19.8	79.2	110000	10 ADG70184_2	Continuation (3 of
46	19.2	76.8	498	6 AEN92804	Abn92804 Staphyloc
47	19.2	76.8	498	13 ADS02942	Ads02942 Staphyloc
48	19.2	76.8	544	6 ABO25827	Abq25827 OligonucI
49	19.2	76.8	544	6 ABO25826	Abq25826 OligonucI
50	19.2	76.8	612	6 AEO24444	Abq24444 OligonucI
51	19.2	76.8	612	6 AEO24445	Abq24445 OligonucI
52	19.2	76.8	655	4 AAK56535	Aak56535 Human imm
53	19.2	76.8	722	6 ABO35161	Abq35161 OligonucI
54	19.2	76.8	722	6 ABO35160	Abq35160 OligonucI
55	19.2	76.8	730	6 ABO15220	Abq15220 OligonucI
56	19.2	76.8	730	6 ABO15221	Abq15221 OligonucI
57	19.2	76.8	1096	4 AAK69138	Aak69138 Human imm
58	19.2	76.8	1529	8 ADA72856	Ada72856 Rice gene
59	19.2	76.8	1834	6 AAS15007	Aas15007 DNA encod
60	19.2	76.8	2000	11 ACL38225	AcL38225 Rice stre
61	19.2	76.8	2000	11 ACL38226	AcL38226 Rice stre
62	19.2	76.8	2000	12 ADJ40625	Adj40625 Plant cDN
63	19.2	76.8	5666	6 AAS61187	Aas61187 Human gen
64	19.2	76.8	5666	6 ABL49330	Abi49330 Human pol
65	19.2	76.8	5745	6 ABE28375	Abk28375 DNA trans
66	19.2	76.8	5763	6 ABL32182	Abi32182 Human imm
67	19.2	76.8	6127	6 ABL33615	Abi33615 Human imm
68	19.2	76.8	6216	6 ABL33933	Abi33933 Human che
69	19.2	76.8	6216	6 ABL70140	Abi70140 ChemicalI
70	19.2	76.8	6221	4 AAS46503	Aas46503 Tumour su
71	19.2	76.8	6379	4 AAS46348	Aas46348 Tumour su
72	19.2	76.8	8951	6 ABL32795	Abi32795 Human imm
73	19.2	76.8	8959	13 ADS89440	Ads89440 OligonucI
74	19.2	76.8	9859	13 ADS89714	Ads89714 OligonucI
75	19.2	76.8	10144	6 ABL43933	Abi43933 Human imm
76	19.2	76.8	13038	6 ABL33275	Abi33275 Human imm
77	19.2	76.8	14542	6 ABL31234	Abi31234 Signal tr
78	19.2	76.8	14542	6 ABL70191	Abi70191 ChemicalI
79	19.2	76.8	14542	6 AAS61147	Aas61147 Human gen
80	19.2	76.8	15872	4 AAS46519	Aas46519 Tumour su
81	19.2	76.8	16597	14 AEA90533	Aea90533 Escherich
82	19.2	76.8	16724	6 ABL33090	Abi33090 Human imm
83	19.2	76.8	16724	6 ABL34536	Abi34536 Human met
84	19.2	76.8	16724	7 ABL70259	Abi70259 ChemicalI
85	19.2	76.8	16724	7 ADS99797	Ads99797 Bieulphic
86	19.2	76.8	17968	14 ADM10346	Adm10346 Colon pro
87	19.2	76.8	17968	14 ADM10488	Adm10488 Colon pro
88	19.2	76.8	32132	4 AAL35943	Aal35943 Human mus
89	19.2	76.8	32132	4 AAL07060	Aal07060 Human rep
90	19.2	76.8	32132	8 ABE58931	AbE58931 cDNA enco
91	19.2	76.8	32132	12 ADJ39681	Adj39681 Human mus
92	19.2	76.8	32169	4 ABL43076	AbL43076 Genomic s

C	93	19.2	76.8	32169	9	ADB61232
C	94	19.2	76.8	58181	8	AB274619
C	95	19.2	76.8	58181	10	AD2C21010
C	96	19.2	76.8	58181	10	AB268140
C	97	19.2	76.8	110000	6	ABA90521.07
C	98	19.2	76.8	110000	10	AD281169.0
C	99	19.2	76.8	110000	14	AB281175_02
C	100	19.2	76.8	110000	14	AB281175_03
C	101	19.2	76.8	110000	14	AB281175_02
C	102	19.2	76.8	130244	11	ABD32872
C	103	19.2	76.8	131576	11	ACN44890
C	104	19.2	76.8	154465	6	AA228763
C	105	19.2	76.8	158245	6	AA228762
C	106	19.2	76.8	161425	4	AAH02340
C	107	19.2	76.8	162025	4	AAH02339
C	108	19.2	76.8	162025	6	AA228758
C	109	19.2	76.8	162025	6	AA228759
C	110	19.2	76.8	162025	13	AD287598
C	111	19.2	76.8	162025	13	AD287599
C	112	19.2	76.8	162025	13	AD287596
C	113	19.2	76.8	162025	13	AD287596
C	114	19.2	76.8	162025	13	AD287596
C	115	19.2	76.8	162025	13	AD287596
C	116	19.2	76.8	162025	13	AD287596
C	117	19.2	76.8	162025	13	AD287596
C	118	19.2	76.8	162025	13	AD287596
C	119	19.2	76.8	162025	13	AD287596
C	120	19.2	76.8	162025	13	AD287596
C	121	19.2	76.8	162025	13	AD287596
C	122	19.2	76.8	162025	13	AD287596
C	123	19.2	76.8	162025	13	AD287596
C	124	19.2	76.8	162025	13	AD287596
C	125	19.2	76.8	162025	13	AD287596
C	126	19.2	76.8	162025	13	AD287596
C	127	19.2	76.8	162025	13	AD287596
C	128	19.2	76.8	162025	13	AD287596
C	129	19.2	76.8	162025	13	AD287596
C	130	19.2	76.8	162025	13	AD287596
C	131	19.2	76.8	162025	13	AD287596
C	132	19.2	76.8	162025	13	AD287596
C	133	19.2	76.8	162025	13	AD287596
C	134	19.2	76.8	162025	13	AD287596
C	135	19.2	76.8	162025	13	AD287596
C	136	19.2	76.8	162025	13	AD287596
C	137	19.2	76.8	162025	13	AD287596
C	138	19.2	76.8	162025	13	AD287596
C	139	19.2	76.8	162025	13	AD287596
C	140	19.2	76.8	162025	13	AD287596
C	141	19.2	76.8	162025	13	AD287596
C	142	19.2	76.8	162025	13	AD287596
C	143	19.2	76.8	162025	13	AD287596
C	144	19.2	76.8	162025	13	AD287596
C	145	19.2	76.8	162025	13	AD287596
C	146	19.2	76.8	162025	13	AD287596
C	147	19.2	76.8	162025	13	AD287596
C	148	19.2	76.8	162025	13	AD287596
C	149	19.2	76.8	162025	13	AD287596
C	150	19.2	76.8	162025	13	AD287596

ALIGNMENTS

RESULT 1
ACN44850/c
ID ACN44850 standard; DNA; 142318 BP.

AC ACN44850;

DT 18-NOV-2004 (first entry)

DE Human genomic sequence hCG33122.

XX

KW	Cytosolic; carcinoma; lymphoma; cancer; human; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO2003073826-A2.
XX	
PD	12-SEP-2003.
XX	
PF	28-FEB-2003; 2003MO-US006235.
XX	
PR	01-MAR-2002; 2002US-00087192.
XX	
PA	(SAGR-) SAGRES DISCOVERY.
XX	
PI	Morris DW;
XX	
DR	WPI; 2003-328604/31.
XX	
PT	Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX	comprises a nucleotide sequence.
XX	
PS	Claim 1; SEQ ID NO 1504; 0pp; English.
XX	
CC	The present invention relates to novel DNA and protein sequences which
XX	are associated with carcinomas. The sequences are useful for: (i) for
CC	screening drug candidates; (ii) for screening of bioactive agent capable
XX	of binding to carcinoma associated protein (CAP); (iii) for screening of
CC	a bioactive agent capable of modulating the activity of CAP; (iv) for
XX	evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC	carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX	carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC	(x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX	determining Carcinoma Associated (CA) gene copy number. In addition, the
CC	CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX	carcinoma including lymphoma. The present sequence is one such CA coding
CC	sequence. Note: This patent is an equivalent to basic patent
XX	US2002182586A1, for which no sequence data was published
XX	
SO	Sequence 142318 BP; 38833 A; 25825 C; 26759 G; 47076 T; 0 U; 3825 Other;
XX	
QY	Query Match 87.2%; Score 21.8; DB 11; Length 142318;
XX	Best Local Similarity 92.0%; Pred. No. 1.6e+02;
XX	Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DB	121397 AAAAAAAAAATCGCAAAAAATCT 121373
QY	1 AAAAAAAAAATCGCAAAAAATCT 25
XX	
RESULT 2	
ID	AA186707 standard; cDNA; 4068 BP.
XX	
AC	AA186707;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 6767.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
XX	nervous system disorders; arthritis; inflammation; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO200164835-A2.
XX	
PD	07-SEP-2001.
XX	
PF	26-FEB-2001; 2001MO-US004927.
XX	
PR	28-FEB-2000; 2000US-00515126.

```
PR 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AA006776.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 1; SEQ ID NO 6767; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 4068 BP; 1134 A; 1070 C; 1053 G; 811 T; 0 U; 0 Other;
XX
XX Query Match 81.6%; Score 20.4; DB 4; Length 4068;
XX Best Local Similarity 95.5%; Pred. No. 4.4e+02;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAATCGCAGACAAA 22
XX |||||||||
XX DB 536 AAAAAAAAAATCGCAGACAAA 557
XX
XX RESULT 3
XX AD280551/C
XX ID AD280551 standard; cDNA; 7300 BP.
XX
XX AC AD280551;
XX
XX DT 14-JUL-2005 (first entry)
XX
XX DE SH3-domain binding protein 2 (SH3BP2) cDNA.
XX
XX KW cardiant; vasotropic; gene therapy; diagnosis; prognosis;
XX KW gene expression; coronary artery disease; cardiant; vasotropic;
XX KW cardiovascular disease; ds; SH3BP2.
XX
XX OS Homo sapiens.
XX
XX WO2005040422-A2.
XX
XX EN 06-MAY-2005.
XX
XX PD 15-OCT-2004; 2004WO-BP011651.
XX
XX PF 16-OCT-2003; 2003US-0511784P.
XX
XX PR 27-MAY-2004; 2004US-0574818P.
XX
XX XX
XX PA (NOVS ) NOVARTIS AG.
XX PA (NOVS ) NOVARTIS PHARMA GMBH.
XX
XX PI Chibout S, Graess P, Vonderscher J;
XX
XX WPI; 2005-355863/36.
XX
XX DR
XX
XX PT Identifying or predicting (the predisposition of) CAD, monitoring a
XX PT subject identified as having CAD before and after treatment or the
```

```
PT progression or severity of CAD by determining specific peptide/gene
XX expression levels.
XX
XX Claim 7; SEQ ID NO 17; 136pp; English.
XX
XX The invention describes a method of identifying or predicting (the
XX CC predisposition of) coronary artery disease (CAD), monitoring a subject
XX CC identified as having CAD before and after treatment or the progression or
XX CC severity of CAD comprising determining the level of one or more peptides
XX CC or gene expression of at least one gene given in the specification in a
XX CC subject to provide a first value and in a control or reference standard
XX CC to provide a second value, and comparing whether there is a difference
XX CC between the first value and second value. Also described are: screening
XX CC candidate agents for use in treatment of CAD; treating or preventing CAD;
XX CC manufacture of a medicament for the treatment or prevention of CAD
XX CC comprising an agent that can induce a decrease in the level of gene
XX CC expression, synthesis, or activity of at least one gene or gene
XX CC expression products; and a kit for the identifying or predicting the
XX CC predisposition CAD in a subject comprising: instructions for determining
XX CC the peptide level and/or level of gene expression; control or reference
XX CC standard peptide level and/or level of gene expression from a normal
XX CC subject(s) without CAD for at least one gene or peptide cited above. A
XX CC substance comprising an agent that can induce a decrease in the level of
XX CC gene expression, synthesis, or activity of at least one gene or gene
XX CC expression product or the level of at least one disease over control
XX CC and/or predominant in disease peptide and/or induce an increase in the
XX CC level of at least one control over disease and/or predominant in control
XX CC peptide and/or a decrease in gene expression, synthesis, or activity of
XX CC at least one gene or gene expression product is useful for manufacturing
XX CC a medicament for the treatment or prevention of CAD. The method is useful
XX CC for identifying or predicting (the predisposition of) CAD, monitoring a
XX CC subject identified as having CAD before and after treatment or monitoring
XX CC the progression or severity of CAD, thus treating CAD. This sequence
XX CC represents a PTEN induced putative kinase 1 polynucleotide, a highly
XX CC predictive gene used to determine the severity of coronary artery
XX CC disease.
XX
XX SQ Sequence 7300 BP; 1478 A; 2074 C; 1983 G; 1765 T; 0 U; 0 Other;
XX
XX Query Match 81.6%; Score 20.4; DB 14; Length 7300;
XX Best Local Similarity 95.5%; Pred. No. 4.5e+02;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAATCGCAGACAAA 22
XX |||||||||
XX DB 6764 AAAAAAAAAATCGCAGACAAA 6743
XX
XX RESULT 4
XX ADCT1368/C
XX ID ADCT1368 standard; cDNA; 13388 BP.
XX
XX AC ADCT1368;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE Human colon specific cDNA sequence DEX0235_64 (SeqID 64).
XX
XX KW human; gene; ss; neoplastic colorectal; colon cancer;
XX KW non-cancerous disease; gene therapy; transgenic; DEX0235_64.
XX
XX OS Homo sapiens.
XX
XX WO2003020934-A1.
XX
XX PN 13-MAR-2003.
XX
XX PD 29-AUG-2002; 2002WO-US027737.
XX
XX PF 31-AUG-2001; 2001US-0316258P.
XX
XX PR
XX
XX PA (DIAD-) DIADEXUS INC.
XX
```

PI Sun Y, Liu C, Ghosh MG;
XX WPI; 2003-300891/29.
XX P-PsDB; ADC71301.
XX
PT Novel colon specific polypeptides and nucleic acids, useful for
PT identifying, diagnosing, monitoring, staging, imaging and treating colon
PT cancer and non-cancerous disease states in colon tissue.
XX
XX Claim 1; SEQ ID NO 64; 262pp; English.
XX
CC This invention relates to novel nucleic acid molecules and the encoded
CC polypeptides, which are present in normal and neoplastic colorectal
CC cells. Specifically, it refers to antibodies of these colon specific
CC polypeptides, as well as antagonists and agonists thereof that can be
CC used to treat colon cancer and also non-cancerous diseases states of the
CC colon. The present invention describes methods useful for the diagnosis
CC and monitoring of colon cancer metastases in a patient, by determining
CC the concentration of these colon specific proteins in a patient sample.
CC Furthermore, they are also used for gene therapy purposes, the production
CC of transgenic animals and cells, as well as producing engineered colon
CC tissue for treatment and research. This polynucleotide sequence is a
CC human colon specific cDNA sequence of the invention.
XX
SQ Sequence 13388 BP; 3129 A; 3503 C; 3340 G; 3414 T; 0 U; 2 Other;
Query Match 81.6%; Score 20.4; DB 10; Length 13388;
Best Local Similarity 95.5%; Pred. No. 4.6e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCCGACACAAA 22
Db 6765 AAAAAAAAAATCCGACACAAA 6744
RESULT 5
ADK70082
ID ADK70082 standard; DNA; 22970 BP.
XX
XX AC ADK70082;
XX
XX DT 06-MAY-2004 (first entry)
XX
XX DE Mutant human SH3 binding protein 2 genomic sequence.
XX
XX KW ds; gene; osteopathic; cytostatic; gene therapy; SH3-binding protein;
XX SH3BP2; mutation; diagnosis; bone homeostasis; cherubism; bone tumor.
XX
XX OS Homo sapiens.
XX OS Synthetic.
XX
XX PN WO2003025197-A2.
XX
XX PD 27-MAR-2003.
XX
XX PF 01-FEB-2002; 2002WO-US019164.
XX
XX PR 02-FEB-2001; 2001US-0266129P.
XX
XX PA (HARD) HARVARD COLLEGE.
XX PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
XX
XX PI Tiziani V, Reichenberger E, Ueki Y, Olsen BR;
XX
XX WPI; 2003-371820/35.
XX
XX PT New mutant SH3BP2 nucleic acid molecule or polypeptide, useful for
XX diagnosing and treating disorders of bone homeostasis, such as cherubism
XX or bone tumor.
XX
XX PS Claim 1; SEQ ID NO 26; 70pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule comprising a

CC mutant SH3-binding protein and its encoding DNA, nucleic acid molecules
CC which hybridize under stringent conditions to it, a nucleic acid
CC molecules that differ from it in codon sequence due to the degeneracy of
CC the genetic code; and complements of these, provided that the nucleic
CC acid molecule is not a human 'WT' SH3BP2 full length sequence. The mutant
CC is especially selected from a genetic mutant domain; 'H' family mutation;
CC 'K' family mutation; 'A,B' family mutation; 'C, F, U, M, O' family
CC mutation; 'L' family mutation; 'G' family mutation; 'N' family mutation
CC or genomic mutant SH3BP2), any of the mutant SH3BP2 exon 9 sequences and
CC a genomic mutant SH3BP2 nucleic acid molecule. The nucleic acid molecule
CC and polypeptide are useful for diagnosing and treating disorders of bone
CC homeostasis, such as cherubism or a bone tumor. This sequence corresponds
CC to the mutant genomic SH3-binding protein DNA sequence.
XX
SQ Sequence 22970 BP; 5407 A; 6744 C; 6741 G; 4069 T; 0 U; 9 Other;
Query Match 81.6%; Score 20.4; DB 10; Length 22970;
Best Local Similarity 95.5%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AAAAAAAAAATCCGACACAAA 22
Db 269 AAAAAAAAAATCCGACACAAA 290
RESULT 6
ADK70081
ID ADK70081 standard; DNA; 22970 BP.
XX
XX AC ADK70081;
XX
XX DT 06-MAY-2004 (first entry)
XX
XX DE Wild type human SH3 binding protein 2 genomic sequence.
XX
XX KW ds; gene; osteopathic; cytostatic; gene therapy; SH3-binding protein;
XX SH3BP2; mutation; diagnosis; bone homeostasis; cherubism; bone tumor.
XX
XX OS Homo sapiens.
XX
XX PN WO2003025197-A2.
XX
XX PD 27-MAR-2003.
XX
XX PF 01-FEB-2002; 2002WO-US019164.
XX
XX PR 02-FEB-2001; 2001US-0266129P.
XX
XX PA (HARD) HARVARD COLLEGE.
XX PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
XX
XX PI Tiziani V, Reichenberger E, Ueki Y, Olsen BR;
XX
XX WPI; 2003-371820/35.
XX
XX PT New mutant SH3BP2 nucleic acid molecule or polypeptide, useful for
XX diagnosing and treating disorders of bone homeostasis, such as cherubism
XX or bone tumor.
XX
XX PS Disclosure; SEQ ID NO 25; 70pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule comprising a
XX mutant SH3-binding protein and its encoding DNA, nucleic acid molecules
XX which hybridize under stringent conditions to it, a nucleic acid
XX molecules that differ from it in codon sequence due to the degeneracy of
XX the genetic code; and complements of these, provided that the nucleic
XX acid molecule is not a human 'WT' SH3BP2 full length sequence. The mutant
XX is especially selected from a genetic mutant domain; 'H' family mutation;
XX 'K' family mutation; 'A,B' family mutation; 'C, F, U, M, O' family
XX mutation; 'L' family mutation; 'G' family mutation; 'N' family mutation
XX or genomic mutant SH3BP2), any of the mutant SH3BP2 exon 9 sequences and
XX a genomic mutant SH3BP2 nucleic acid molecule. The nucleic acid molecule
XX and polypeptide are useful for diagnosing and treating disorders of bone

CC homeostasis, such as chernobism or a bone tumor. This sequence corresponds
CC to the wild type genomic SH3-binding protein DNA sequence.
SQ Sequence 22970 BP; 5407 A; 6748 C; 6745 G; 4070 T; 0 U; 0 Other;
Query Match 81.6%; Score 20.4; DB 10; Length 22970;
Best Local Similarity 95.5%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAGCAAA 22
DB 269 AAAAAAAAAATCGCAGCAAA 290
RESULT 7
ADZ13116/c
ID ADZ13116 standard; DNA; 241748 BP.
XX
AC ADZ13116;
XX
DT 16-JUN-2005 (first entry)
XX
DE Murine cancer-associated genomic DNA #54.
XX
KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasia;
KM cytostatic; gene; ds.
XX
OS Mus sp.
XX
PN WO2005031001-A2.
XX
PD 07-APR-2005.
XX
PF 23-SEP-2004; 2004WO-US031617.
XX
PR 23-SEP-2003; 2003US-00669920.
XX
PA (CHIR) CHIRON CORP.
XX
PI Morris DW, Malandro MS;
XX
PS WPI; 2005-273395/28.
XX
DR Nucleic acid array useful for detecting cancer associated nucleic acid,
PT comprises two or more nucleic acid probes.
XX
PI Nucleic acid array useful for detecting cancer associated nucleic acid,
PT comprises two or more nucleic acid probes.
XX
PS Disclosure; SEQ ID NO 636; 198pp; English.
XX
CC The invention relates to a nucleic acid array for detecting a cancer
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC The invention also relates to a peptide array comprising two or more
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC that binds to a polypeptide, an isolated antibody or its fragment which
CC binds to a polypeptide, which is prepared by immunizing a host animal
CC with a composition comprising the polypeptide or its antigen binding
CC fragment and collecting cells from the host expressing antibodies against
CC the antigen or its antigen binding fragment, a composition comprising the
CC antibody and a carrier, a method of screening for anticancer activity, a
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC method of treating cancer and a method of inhibiting expression of a CA
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC nucleic acids. The antibody is useful for detecting the presence or
CC absence of cancer cells in an individual which involves contacting cells
CC from the individual with the antibody and detecting a complex of a CA
CC protein from the cancer cells and the antibody, where the detection of
CC the complex correlates with the presence of cancer cells in the
CC individual. The composition is useful for inhibiting growth of cancer
CC cells in an individual or for delivering a therapeutic agent to cancer
CC cells in an individual. The invention is also useful for diagnosing
CC cancer, for treating cancer and for inhibiting expression of a CA gene in
CC a cell. This sequence represents murine cancer-associated genomic DNA of
XX the invention.

SQ Sequence 241748 BP; 59664 A; 56662 C; 57444 G; 61211 T; 0 U; 6767 Other;
Query Match 81.6%; Score 20.4; DB 14; Length 241748;
Best Local Similarity 95.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGCAGCAAA 22
DB 160343 AAAAAAAAAATCGCAGCAAA 160322
RESULT 8
AAC49529/c
ID AAC49529 standard; DNA; 1532 BP.
XX
AC AAC49529;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 61493.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0136239P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137502P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0143390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144332P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0153707P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159332P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 80.8%; Score 20.2; DB 3; Length 1532;
Best Local Similarity 80.0%; Pred. No. 5e+02; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 AAAAAAAAAATCGACGACCAATCT 25

Db 1311 AAAAAAAAACTTGCAGAGAAATCT 1287
RESULT 9
ID AAC42454/c
AAC42454 standard; DNA; 1536 BP.
XX
AC AAC42454;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 35619.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200SEP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130499P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137528P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138647P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145919P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148655P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.

PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155137P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0158293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160747P.
PR	21-OCT-1999;	99US-0160776P.
PR	21-OCT-1999;	99US-0160776P.
PR	21-OCT-1999;	99US-0160779P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160988P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	28-OCT-1999;	99US-0161935P.
PR	29-OCT-1999;	99US-0162142P.

```
Query Match Similarity      80.8%; Score 20.2; DB 3; Length 1536;
Best Local Similarity       88.0%; Pred.No. 3e+02;
Matches    22; Conservative   0; Mismatches     3; Indels    0; Gaps    0;

QY      1 AAAAAAAAAATCGCAGCAAAATCT 25
          ||||| | | | | | | | | |
Db      1315 AAAAAAAAAACTTGACAGGAATCT 1291

RESULT 10
ABL28990 standard; DNA; 3875 BP.
```

XX	ABU28990;
AC	
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 38443.
XX	
XX	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ds.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US009231.
XX	
PR	23-MAR-2000; 2000US-0191637P.
XX	
PR	11-JUL-2000; 2000US-00614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
XX	genes from Drosophila and for elucidating cell signaling and cell-cell
PT	interactions.
XX	
PS	Claim 1; SEQ ID NO 38443; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signaling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC	sequences (ABU01840-ABU16175) and the encoded proteins (AB857737-
CC	AB872072). The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX	
SQ	Sequence 3875 BP; 1056 A; 818 C; 834 G; 1167 T; 0 U; 0 Other;
XX	
Query Match	80.8%; Score 20.2; DB 4; Length 3875;
Best Local Similarity	88.0%; Pred. No. 5.2e+02;
Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
QY	1 AAAAAAAAAATCGAGCAAAATCT 25
DB	1762 AAAAAAAAAATTGACAGCAAAATT 1786
XX	
RESULT 11	
ABU92315/C	
ID	ABU92315 standard; DNA; 6162 BP.
XX	
AC	ABU92315;
XX	
DT	01-JUL-2002 (first entry)
XX	
DE	Chemically treated DNA repair gene fragment complementary to#62.
XX	
KW	DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L3; PMS2;
XX	L4; PMS2L5; PMS2L6; MGMT; NUDT1; TDG; INPBL; RFC4; DITIL; FANCB;
KW	XRCB8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;
KW	Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
KW	trichiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.
XX	
OS	Unidentified.
XX	

PN WO200181622-A2.
XX
PD 01-NOV-2001.
XX
XX 06-APR-2001; 2001WO-EP003972.
PF
XX 06-APR-2000; 2000DE-01019058.
PR
XX 07-APR-2000; 2000DE-01019173.
PR
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPICGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K,
XX
DR WPI; 2002-034446/04.
XX
XX New nucleic acid derived from genes associated with DNA repair, useful
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT cytosine methylation.
XX
PS Claim 1; SEQ ID NO 124; 25pp + Sequence Listing; English.
XX
XX The invention relates to nucleic acids containing a sequence of at least
CC 18 nucleotides of chemically treated DNA of genes associated with DNA
CC repair, and their complements. The invention also relates to nucleic
CC acids comprising at least 18 base pairs of the chemically pretreated DNA
CC of genes associated with DNA repair selected from PMS2L1, PMS2L2,
CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, MSH1, TDC, INP1L,
CC RFC4, DIT1L, FANCB, or XRCC8. Nucleic acids of the invention and related
CC oligomers, are useful for diagnosis of diseases associated with gene
CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC Cockayne syndrome, Niemann breakage syndrome or Werner syndrome,
CC immunodeficiency, trichiodystrophy, Fanconi's anaemia, solid tumours
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records AB192192-AB192335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: The sequence data for this patent is
CC not represented in the specification, but is based on sequence
CC information supplied by the European Patent Office
XX
SQ Sequence 6162 BP; 1341 A; 178 C; 1476 G; 3167 T; 0 U; 0 Other;
XX
Query Match 80.8%; Score 20.2; DB 6; Length 6162;
Best Local Similarity 88.0%; Pred No. 5.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGACGAAATCT 25
DB 3579 AAAAAAAAAACGAAAAAAAAATCT 3555
RESULT 12
ABK43074/C
ID ABK43074 standard; DNA; 6226 BP.
XX
XX ABK43074;
AC
XX 21-MAY-2002 (first entry)
DT
XX
XX Genomic sequence #973 encoding novel human connective tissue polypeptide.
DE
XX
XX Human; connective tissue related disorder; cancer; gene therapy;
KM cytosstatic; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200155343-A1.
PN
XX
XX 02-AUG-2001.
PD
XX

PF 17-JAN-2001; 2001WO-US001322.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR
XX 04-FEB-2000; 2000US-0180628P.
PR
XX 24-FEB-2000; 2000US-0184664P.
PR
XX 02-MAR-2000; 2000US-0186350P.
PR
XX 16-MAR-2000; 2000US-0189874P.
PR
XX 17-MAR-2000; 2000US-0190076P.
PR
XX 18-APR-2000; 2000US-0198123P.
PR
XX 19-MAY-2000; 2000US-0205515P.
PR
XX 07-JUN-2000; 2000US-0209467P.
PR
XX 28-JUN-2000; 2000US-0214886P.
PR
XX 30-JUN-2000; 2000US-0215135P.
PR
XX 07-JUL-2000; 2000US-0216880P.
PR
XX 11-JUL-2000; 2000US-0217487P.
PR
XX 14-JUL-2000; 2000US-0217496P.
PR
XX 26-JUL-2000; 2000US-0218290P.
PR
XX 26-JUL-2000; 2000US-0220963P.
PR
XX 26-JUL-2000; 2000US-0220964P.
PR
XX 14-AUG-2000; 2000US-0224518P.
PR
XX 14-AUG-2000; 2000US-0225213P.
PR
XX 14-AUG-2000; 2000US-0225214P.
PR
XX 14-AUG-2000; 2000US-0225266P.
PR
XX 14-AUG-2000; 2000US-0225267P.
PR
XX 14-AUG-2000; 2000US-0225268P.
PR
XX 14-AUG-2000; 2000US-0225270P.
PR
XX 14-AUG-2000; 2000US-0225447P.
PR
XX 14-AUG-2000; 2000US-0225757P.
PR
XX 14-AUG-2000; 2000US-0225758P.
PR
XX 14-AUG-2000; 2000US-0225759P.
PR
XX 18-AUG-2000; 2000US-0226279P.
PR
XX 22-AUG-2000; 2000US-0226681P.
PR
XX 22-AUG-2000; 2000US-0226682P.
PR
XX 22-AUG-2000; 2000US-0227182P.
PR
XX 23-AUG-2000; 2000US-0227099P.
PR
XX 30-AUG-2000; 2000US-0228924P.
PR
XX 01-SEP-2000; 2000US-0229287P.
PR
XX 01-SEP-2000; 2000US-0229343P.
PR
XX 01-SEP-2000; 2000US-0229344P.
PR
XX 01-SEP-2000; 2000US-0229345P.
PR
XX 05-SEP-2000; 2000US-0229309P.
PR
XX 05-SEP-2000; 2000US-0229513P.
PR
XX 06-SEP-2000; 2000US-0230437P.
PR
XX 06-SEP-2000; 2000US-0230438P.
PR
XX 08-SEP-2000; 2000US-0231242P.
PR
XX 08-SEP-2000; 2000US-0231243P.
PR
XX 08-SEP-2000; 2000US-0231244P.
PR
XX 08-SEP-2000; 2000US-0231413P.
PR
XX 08-SEP-2000; 2000US-0231414P.
PR
XX 08-SEP-2000; 2000US-0233080P.
PR
XX 08-SEP-2000; 2000US-0233081P.
PR
XX 12-SEP-2000; 2000US-0233968P.
PR
XX 14-SEP-2000; 2000US-0233977P.
PR
XX 14-SEP-2000; 2000US-0233988P.
PR
XX 14-SEP-2000; 2000US-0233999P.
PR
XX 14-SEP-2000; 2000US-0234000P.
PR
XX 14-SEP-2000; 2000US-0234001P.
PR
XX 14-SEP-2000; 2000US-0233063P.
PR
XX 14-SEP-2000; 2000US-0233064P.
PR
XX 14-SEP-2000; 2000US-0233065P.
PR
XX 21-SEP-2000; 2000US-0234223P.
PR
XX 21-SEP-2000; 2000US-0234274P.
PR
XX 25-SEP-2000; 2000US-0234997P.
PR
XX 25-SEP-2000; 2000US-0234998P.
PR
XX 26-SEP-2000; 2000US-0234984P.
PR
XX 27-SEP-2000; 2000US-0235834P.
PR
XX 27-SEP-2000; 2000US-0235835P.
PR
XX 29-SEP-2000; 2000US-0236327P.
PR
XX 29-SEP-2000; 2000US-0236367P.
PR
XX 29-SEP-2000; 2000US-0236368P.
PR
XX 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P-
 PR 02-OCT-2000; 2000US-0236802P-
 PR 02-OCT-2000; 2000US-0237037P-
 PR 02-OCT-2000; 2000US-0237038P-
 PR 02-OCT-2000; 2000US-0237039P-
 PR 02-OCT-2000; 2000US-0237040P-
 PR 13-OCT-2000; 2000US-0239335P-
 PR 13-OCT-2000; 2000US-0239337P-
 PR 20-OCT-2000; 2000US-0240960P-
 PR 20-OCT-2000; 2000US-0241221P-
 PR 20-OCT-2000; 2000US-0241785P-
 PR 20-OCT-2000; 2000US-0241786P-
 PR 20-OCT-2000; 2000US-0241787P-
 PR 20-OCT-2000; 2000US-0241808P-
 PR 20-OCT-2000; 2000US-0241809P-
 PR 20-OCT-2000; 2000US-0241826P-
 PR 01-NOV-2000; 2000US-0244617P-
 PR 08-NOV-2000; 2000US-0246474P-
 PR 08-NOV-2000; 2000US-0246475P-
 PR 08-NOV-2000; 2000US-0246476P-
 PR 08-NOV-2000; 2000US-0246477P-
 PR 08-NOV-2000; 2000US-0246523P-
 PR 08-NOV-2000; 2000US-0246524P-
 PR 08-NOV-2000; 2000US-0246525P-
 PR 08-NOV-2000; 2000US-0246526P-
 PR 08-NOV-2000; 2000US-0246527P-
 PR 08-NOV-2000; 2000US-0246528P-
 PR 08-NOV-2000; 2000US-0246532P-
 PR 08-NOV-2000; 2000US-0246609P-
 PR 08-NOV-2000; 2000US-0246610P-
 PR 08-NOV-2000; 2000US-0246611P-
 PR 17-NOV-2000; 2000US-0249207P-
 PR 17-NOV-2000; 2000US-0249208P-
 PR 17-NOV-2000; 2000US-0249209P-
 PR 17-NOV-2000; 2000US-0249210P-
 PR 17-NOV-2000; 2000US-0249211P-
 PR 17-NOV-2000; 2000US-0249212P-
 PR 17-NOV-2000; 2000US-0249213P-
 PR 17-NOV-2000; 2000US-0249214P-
 PR 17-NOV-2000; 2000US-0249215P-
 PR 17-NOV-2000; 2000US-0249216P-
 PR 17-NOV-2000; 2000US-0249217P-
 PR 17-NOV-2000; 2000US-0249218P-
 PR 17-NOV-2000; 2000US-0249244P-
 PR 17-NOV-2000; 2000US-0249245P-
 PR 17-NOV-2000; 2000US-0249264P-
 PR 17-NOV-2000; 2000US-0249265P-
 PR 17-NOV-2000; 2000US-0249297P-
 PR 17-NOV-2000; 2000US-0249299P-
 PR 17-NOV-2000; 2000US-0249300P-
 PR 01-DEC-2000; 2000US-0250160P-
 PR 01-DEC-2000; 2000US-0250391P-
 PR 05-DEC-2000; 2000US-0251030P-
 PR 05-DEC-2000; 2000US-0251988P-
 PR 05-DEC-2000; 2000US-0256719P-
 PR 06-DEC-2000; 2000US-0256719P-
 PR 08-DEC-2000; 2000US-0251856P-
 PR 08-DEC-2000; 2000US-0251868P-
 PR 08-DEC-2000; 2000US-0251869P-
 PR 08-DEC-2000; 2000US-0251899P-
 PR 08-DEC-2000; 2000US-0251990P-
 PR 11-DEC-2000; 2000US-0256097P-
 PR 05-JAN-2001; 2001US-0259678P-
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Barash SC, Ruben SM;
 PI
 XX WPI; 2001-565190/63.
 DR
 XX Nucleic acid encoding novel connective tissue associated polypeptides,
 PT

PT used in diagnosing, preventing, treating or ameliorating a disorder such
 as cancer or rheumatoid arthritis.
 PR
 XX
 XX Disclosure; SEQ ID NO 1961; 673bp; English.
 PS
 XX The present invention relates to the isolation of novel human connective
 CC tissue related polypeptides (A086435-A086923) and the polynucleotide
 CC (cDNA and genomic) sequences encoding them. The sequences of the
 CC invention are useful in the diagnosis, treatment, prevention and/or
 CC prognosis of diseases associated with connective tissue(s), including
 CC cancer. The polynucleotide sequences of the invention are also useful in
 CC gene therapy. ABK42102-ABK43116 represent genomic sequences encoding the
 CC novel human connective tissue related polypeptides. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 6226 BP; 1500 A; 1552 C; 1518 G; 1656 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 4; Length 6226;
 Best Local Similarity 88.0%; Pred. No. 5.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAATCGACAGCAATCT 25
 Db 660 AAAAAAAAAATCGCAAAATATCT 636
 RESULT 13
 ADB61230/c
 ID ADB61230 standard; DNA; 6226 BP.
 XX
 XX ADB61230;
 AC
 XX
 DT 04-DEC-2003 (first entry)
 XX
 XX Connective tissue related genomic DNA #973.
 DE
 XX
 KW cytoslatic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;
 KW antiarteriosclerotic; immunosuppressive; antirheumatic; antiarthritic;
 KW antiinflammatory; antiallergic; antiaschemic; dermatological;
 KW nephrotoxic; virucide; fungicide; antibacterial; antiparasitic;
 KW gene therapy; ds; connective tissues disorder; rheumatoid arthritis;
 KW systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;
 KW cancer metastasis; neoplasia; leukemia; neurodegenerative disorder;
 KW Alzheimer's disease; Parkinson's disease; cardiovascular disease;
 KW atherosclerosis; myocarditis; cardiopulmonary bypass complication;
 KW autoimmune disease; multiple sclerosis; allergic reaction; asthma;
 KW rhinitis; eczema; inflammatory condition; Crohn's disease; nephritis;
 KW gastrointestinal disorder; inflammatory bowel disease;
 KW organ transplant rejection; immune system disorder; Bruton's disease;
 KW X-linked lymphoproliferative syndrome;
 KW B-cell lymphoproliferative disorder; HIV; AIDS; infection;
 KW chromosome identification; chromosome mapping;
 KW connective tissue related polynucleotide; gene; ds.
 OS Homo sapiens.
 XX
 XX US2003054375-A1.
 PN
 XX
 XX 20-MAR-2003.
 PD
 XX
 XX 07-MAR-2002; 2002US-00092154.
 PF
 XX
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR

```
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227099P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246533P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251038P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764847.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Ruben SM, Barash SC;
WPI, 2003-634869/60.
P-PSDB; ADB59748.

New connective tissue-related polypeptides and polynucleotides, useful
for treating, preventing and/or prognosing e.g. disorders of connective
tissue, (e.g. rheumatoid arthritis), cancers, cancer metastases and/or
neoplasias.

Disclosure; SEQ ID NO 1961; 248pp; English.

The invention describes an isolated nucleic acid molecule (I), which
comprises a sequence that is at least 95 % identical to a connective
tissue-related polynucleotide encoding connective tissue antigens (CTA).
```

CC The polypeptide or polynucleotide is useful for preventing, treating, or
 CC ameliorating medical conditions in a mammal. The connective tissue
 CC polypeptides, polynucleotides and antibodies are particularly useful for
 CC treating, preventing and/or diagnosing disorders of connective tissues
 CC (e.g. rheumatoid arthritis, discoid and systemic lupus erythematosus,
 CC scleroderma, or Sjogren's syndrome), cancer, cancer metastases and/or
 CC neoplasias (e.g. leukemia), neurodegenerative disorders (e.g.
 CC Alzheimer's disease, or Parkinson's disease), cardiovascular diseases
 CC (e.g. atherosclerosis, myocarditis or cardiopulmonary bypass
 CC complications), autoimmune diseases (e.g. systemic lupus erythematosus,
 CC rheumatoid arthritis), or multiple sclerosis), allergic reactions (e.g.

Query Match 80.8%; Score 20.2; DB 9; Length 6226;
 Best Local Similarity 88.0%; Pred. No. 5.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
 Db 660 AAAAAAAAAATCGCAAAAATATCT 636

RESULT 14
 ABL92234/C
 ID ABL92234 standard; DNA; 7145 BP.

AC ABL92234;

DT 01-JUL-2002 (first entry)

DE Chemically treated DNA repair gene fragment#22.

XX DNA repair; cytosine methylation; PMS2L1; PMS2L2; PMS2L3; PMS2;
 XX L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RRC4; DIT1L; FANCB;
 XX XRC8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;
 XX Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
 XX trichothiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.

OS Unidentified.

XX WO200181622-A2.

PN 01-NOV-2001.

PF 06-APR-2001; 2001WO-EP003972.

PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-034446/04.

PT New nucleic acid derived from genes associated with DNA repair, useful
 PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
 PT cytosine methylation.

PS Claim 1; SEQ ID NO 43; 25pp + Sequence Listing; English.

XX The invention relates to nucleic acids containing a sequence of at least
 CC 18 nucleotides of chemically treated DNA of genes associated with DNA
 CC repair, and their complements. The invention also relates to nucleic
 CC acids comprising at least 18 base pairs of the chemically pretreated DNA
 CC of genes associated with DNA repair selected from PMS2L1, PMS2L2,
 CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,
 CC RRC4, DIT1L, FANCB, or XRC8. Nucleic acids of the invention and related
 CC oligomers, are useful for diagnosis of diseases associated with gene
 CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
 CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
 CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours

CC and cancer, particularly by determining status of cytosine methylation
 CC and/or by detecting single-nucleotide polymorphisms. Determination of
 CC individual methylation patterns may allow development of individualised
 CC therapies. The sequences given in records ABL92192-ABL92335 represent
 CC chemically pre-treated DNA fragments from genes associated with DNA
 CC repair, and their complements. Note: The sequence data for this patent is
 CC not represented in the specification, but is based on sequence
 CC information supplied by the European Patent Office

Sequence 7145 BP; 1896 A; 177 C; 1529 G; 3543 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 6; Length 7145;
 Best Local Similarity 88.0%; Pred. No. 5.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGAGCAAAATCT 25
 Db 1073 AAAAAAAAAATCGAAAAAAATCT 1049

RESULT 15
 AAD22321/C
 ID AAD22321 standard; DNA; 7145 BP.

AC AAD22321;

DT 12-FEB-2002 (first entry)

DE Chemically treated human genomic DNA #11 associated with DNA adducts.

XX DNA adduct; peptide nucleic acid; PNA; cytosine methylation;
 XX gene therapy; tumour; cancer; human; ds.

XX Homo sapiens.

XX WO200177378-A2.

PN 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP004015.

PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-010923/01.

PT Novel nucleic acid comprising sequence of a segment of chemically
 PT pretreated DNA of genes associated with DNA adduct, useful for diagnosis
 PT and therapy of solid tumors and cancer.

PS Claim 1; Page 52-53; 97pp; English.

XX The invention relates to chemically modified DNA of genes associated with
 CC DNA adducts, oligonucleotides and/or peptide nucleic acid (PNA) oligomers
 CC for detecting cytosine methylations as well as method for ascertaining
 CC genetic and/or epigenetic parameters of genes associated with DNA
 CC adducts. Oligomers of the invention coupled to a solid phase is useful
 CC for manufacturing an arrangement of different oligomers (array) fixed to
 CC a carrier material for analysing diseases associated with the methylation
 CC state of the CpG dinucleotides of chemically pretreated DNA of genes
 CC associated with DNA adduct. They are also useful for ascertaining genetic
 CC and/or epigenetic parameters for the diagnosis and/or therapy of existing
 CC diseases or the predisposition to specific diseases by analysing cytosine
 CC methylations. Sequences of the invention are useful for diagnosis and
 CC therapy of solid tumours and cancers. They are also useful in gene
 CC therapy. The present sequence is chemically pretreated human genomic DNA
 CC associated DNA adducts

XX Sequence 7145 BP; 1896 A; 177 C; 1529 G; 3543 T; 0 U; 0 Other;
SQ
Query Match 80.8%; Score 20.2; DB 6; Length 7145;
Best Local Similarity 88.0%; Pred. No. 5.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 1073 AAAAAAAAAATCGAAAAAAAAATCT 1049
RESULT 16
ABL32095/C
ID ABL32095 standard; DNA; 7928 BP.
XX
AC ABL32095;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 68.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiserotoclerotic; antianaemic; cyrostatic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW db.
XX
OS Homo sapiens.
XX
PN WO20020928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR MPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 68; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
SQ Sequence 7928 BP; 1959 A; 198 C; 2043 G; 3728 T; 0 U; 0 Other;
Qy Query Match 80.8%; Score 20.2; DB 6; Length 7928;
Best Local Similarity 88.0%; Pred. No. 5.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 3542 AAAAAAAAAATCGCAAAATTAATCT 3518

RESULT 17
ABK31169/C
ID ABK31169 standard; DNA; 7928 BP.
XX
AC ABK31169;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified complementary DNA #6.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cyostatic; mutant; db.
XX
OS Homo sapiens.
XX
OS Synthetic.
XX
PN WO20020926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR MPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 12; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
SQ Sequence 7928 BP; 1959 A; 198 C; 2043 G; 3728 T; 0 U; 0 Other;
Qy Query Match 80.8%; Score 20.2; DB 6; Length 7928;
Best Local Similarity 88.0%; Pred. No. 5.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 3542 AAAAAAAAAATCGCAAAATTAATCT 3518
RESULT 18
ABL70124/C
ID ABL70124 standard; DNA; 7928 BP.

```

XX ABU70124;
XX
XX 01-JUN-2002 (first entry)
XX
XX Chemically treated cell signalling DNA sequence complementary to#7.
XX
XX Cell signalling; cytosine methylation; cell signalling disease; cancer;
XX tumour; cytostatic; ds.
XX
XX Unidentified.
XX
XX WO200202807-A2.
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-EP007471.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A. Piepenbrock C, Berlin K;
XX
XX WPI; 2002-154758/20.
XX
XX Nucleic acid, useful for diagnosis and therapy of diseases associated
XX with cell signalling e.g. cancer, comprises chemically modified genomic
XX sequences of genes associated with cell signalling.
XX
XX Claim 1; SEQ ID NO 14; 24pp + Sequence Listing; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of at least
XX 18 bases of a segment of chemically pretreated DNA of genes associated
XX with cell signalling. The activity of the modified sequences of the
XX invention may be described as cytostatic. The object of the invention is
XX to provide the chemically modified DNA of genes associated with cell
XX signalling, as well as oligonucleotides and/or PNA-oligomers for
XX detecting cytosine methylations, as well as a method which is
XX particularly suitable for the diagnosis and/or therapy of genetic and
XX epigenetic parameters of genes associated with cell signalling. The
XX chemically modified DNA provided by the invention is useful for diagnosis
XX and therapy of diseases such as solid tumours and cancer. The sequences
XX given in records ABU7011-ABU7026 represent chemically pre-treated
XX genomic DNA's of genes associated with cell signalling. Note: The
XX sequence data for this patent is not represented in the printed
XX specification, but is based on sequence information supplied by the
XX European Patent Office
XX
XX Sequence 7928 BP; 1959 A; 198 C; 2043 G; 3728 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 6; Length 7928;
XX Best Local Similarity 88.0%; Pred. No. 5.4e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
XX
XX 1 AAAAAAAAAATCGAGCAATCT 25
XX ||||| |||||
XX 3542 AAAAAAAAAATCCCAATTAATCT 3518
XX
XX RESULT 19
XX AAS61056/c
XX ID AAS61056 standard; DNA; 7928 BP.
XX
XX AAS61056;
XX
XX 29-JAN-2002 (first entry)
XX
XX Human gene regulation-associated gene oligonucleotide #11.
XX
XX Human: Gene regulation-associated gene; severe combined immunodeficiency;
XX cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
XX

```

KM	aesthesia; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KM	renal disease; Preeclampsia; cardiac allograft vascular disease;
KM	colorectal cancer; thyroid cancer; oesophageal cancer; de; tumour;
KM	immunostimulant; cadizant; antinflammatory; coagulant; antiasthmatic;
KM	nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX	
OS	Homo sapiens.
PN	WO200177375-A2.
PD	18-OCT-2001.
XX	
PF	06-APR-2001; 2001WO-EP003968.
XX	
PR	06-APR-2000; 2000DE-01019058.
PR	07-JUN-2000; 2000DE-01019173.
PR	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
PA	(EPIG-) EPIGENOMICS AG.
PI	Olek A, Piepenbrock C, Berlin K;
DR	WPI; 2002-017470/02.
XX	
PT	New nucleic acid sequences from chemically modified genes associated with
PT	gene regulation, useful for analyzing cytosine methylations for diagnosis
PT	and therapy of diseases e.g. severe combined immunodeficiency disease.
XX	
PS	Claim 1; SEQ ID NO 12; 26pp; English.
XX	
CC	The invention relates to 224 nucleic acid sequences comprising at least
CC	18 bases of a chemically pretreated gene associated with gene regulation
CC	selected from 43 known genes (or complementary sequences). The chemical
CC	pretreatment converts cytosine bases unmethylated at the 5-position to
CC	uracil or another base with hybridisation behaviour dissimilar to
CC	cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC	oligonucleotides (or sets/arrows) and method are useful in the diagnosis of
CC	diseases (or predisposition to diseases) associated with gene regulation
CC	and in therapy of such diseases, by enabling analysis of the cytosine
CC	methylation patterns of such genes, kits are provided. They are
CC	especially useful in diagnosis and therapy of e.g. severe combined
CC	immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC	and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC	syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC	present sequence is a sequence included in the sequence data for this
CC	specification and is associated with the human gene regulation-associated
CC	genes. Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ	Sequence 7928 BP; 1959 A; 198 C; 2043 G; 3728 T; 0 U; 0 Other;
Query Match:	80.8%; Score 20.2; DB 6; Length 7928;
Best Local Similarity	88.0%; Pred. No. 5.4e+02;
Matches 22; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
Yr	1 AAAAAAAAAATCGACACAATCT 25
Db	3542 AAAAAAAAAATCCCCAATTAAATCT 3518
RESULT 20	
ID	AAK84095/c
XX	AAK84095 standard; DNA; 17677 BP.
AC	AAK84095;
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38907.
KM	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cyostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
XX WC200157182-A2.
PN
XX 09-AUG-2001.
PD
XX 17-JAN-2001; 2001WC-US001354.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0233400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 05-DEC-2000; 2000US-0254799P.
PR 08-DEC-2000; 2000US-0251865P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

```
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rozen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 38907; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM62170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK76594 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM62169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 17677 BP; 3512 A; 5040 C; 4891 G; 4234 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 4; Length 17677;
XX Best local Similarity 88.0%; Pred. No. 5.6e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAATCGAGCAAAATCT 25
XX 1156 AAAAAAAAAATCAGCAAAATCT 1132
XX
XX RESULT 21
XX ID AAK70506/C
XX AAK70506 standard; DNA; 17677 BP.
XX
XX AC AAK70506;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25318.
XX
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX OS Homo sapiens.
XX
XX MO200157182-A2.
XX
XX PV 09-AUG-2001.
XX
XX PD 17-JAN-2001; 2001WO-US001354.
XX
XX PF 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
```

```
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-022681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239393P.
PR 13-OCT-2000; 2000US-0239397P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241281P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
```

PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	01-DEC-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	05-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-025678P.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-483426/52.	
XX		
DR	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
XX	useful for preventing, diagnosing and/or treating cancers and metastasis.	
XX		
XX	Disclosure; SEQ ID NO 25318; 3071pp + Sequence Listing; English.	
PS		
XX	AAM54951 to AAK4702 encode the human immune/hematopoietic antigen (I)	
CC	amino acid sequences given in AAM8210 to AAM91921. (I) have cytosolic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	suppress the patients own production of (I). Additionally, (I)	
CC	may be used to supplement the patients own production of (I).	

CC polynucleotides may be used to produce the secreted (1), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein; (1) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK7694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX

Sequence 17677 BF; 3512 A; 5040 C; 4891 G; 4234 T; 0 U; 0 Other;

Query Match	80.8%	Score 20.2;	DB 4;	length 17677;
Best Local Similarity	88.0%;	Pred. No. 5.6e+02;		
Matches	22;	Conservative	0;	Mismatches 3;
			Indels	0;
			Gaps	0

OY	1	AAAAAAAAAATCGAGACAAATCT	25
Db	1156	AAAAAAAAAATCAGCAAAATCT	1132

RESULT 22	
AAD56081	
ID	AAD56081 standard; DNA; 29105 BP.

DT	07-AUG-2003 (first entry)
XX	
DE	Human MAP3K6 carcinoma associated (CA) gene

KW Carcinoma; gene therapy; carcinoma associated gene; CA; MAP3K6; human;
KW ds.

OS	Homo sapiens.
XX	
PN	W02003035837-A2.

XX	22-OCT-2002; 2002WO-US033835.
PF	
PR	23-OCT-2001; 2001US-00004113.

aa Engelhard EK, Morris DW;
PI
XX
DR WPI; 2003-421412/39.

PT New recombinant nucleic acid and its encoded protein, useful for
PT preparing a composition for diagnosing or treating carcinomas.
XX
PS Claim 1; Page 62-66; 173pp; English.

CC The invention relates to novel sequences which are useful for preparing a
CC composition for diagnosing or treating carcinomas. These sequence are
CC also useful in gene therapy. The present sequence is human MP3K6
CC carcinoma associated (CA) gene. This sequence is used in the invention
XX
Sequence 29105 BR: 5477 A: 7362 C:7988 G: 5611 T: 0 U: 2667 Other:

Query Match	80.8%	Score 20.2;	DB 8;	Length 29105;
Best Local Similarity	88.0%;	Pred. No. 5.7e+02;		
Matches 22; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

```

QY      1 AAAAAAAAAAATCGACACAATCT 25
          |||||  |||||  |||||
Db      28585 AAAAAAAAAAAGCAGACAGATCT 28609

```

RESULT 23
ADA02443
ID ADA02443 standard: DNA: 29105 BP.

```
XX ADA02443;
XX AC
XX DT 06-NOV-2003 (first entry)
XX DE Human MAP3K6 carcinoma associated gene, SEQ ID NO:962.
XX DE
XX DE Human MAP3K6 carcinoma associated gene, SEQ ID NO:962.
XX DE
XX DE Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX DE prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX DE gene; ds.
XX OS
XX OS Homo sapiens.
XX OS
XX OS WO2003057146-A2.
XX OS
XX OS 17-JUL-2003.
XX OS
XX OS 26-DEC-2002; 2002WO-US041414.
XX OS
XX OS 26-DEC-2001; 2001US-00035832.
XX OS
XX OS (SAGR-) SAGRES DISCOVERY.
XX OS
XX OS Morris DW;
XX OS
XX OS WPI; 2003-587068/55.
XX OS
XX PT New recombinant nucleic acid encoding carcinoma associated protein,
XX PT useful for preparing compositions for treating carcinomas.
XX PT
XX PS Claim 1; SEQ ID NO 962; 245bp; English.
XX PS
XX CC The invention relates to recombinant carcinoma associated (CA) nucleic
XX CC acid sequences from mouse and human (ADA01482-ADA0309), and to
XX CC recombinant carcinoma associated proteins (CAP) encoded by them. The
XX CC invention also encompasses expression vectors and host cells comprising a
XX CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
XX CC binds to the protein, and a biochip comprising CA nucleic acid or
XX CC fragments thereof. The sequences of the invention were identified using
XX CC oncogenic retroviruses, which insert into the genome of the host organism
XX CC at random. Many of these do not carry transduced host oncogenes or
XX CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
XX CC direct consequence of the effects of proviral integration into host
XX CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
XX CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
XX CC leukaemia) or a propensity to carcinoma by determination of the sequence
XX CC of a CA gene, or by determination of CA gene expression in particular
XX CC tissues. CA nucleic acids, proteins and antibodies are also useful as
XX CC therapeutic agents and in screening and evaluating drug candidates. The
XX CC present sequence represents a specifically claimed human CA nucleic acid
XX CC sequence of the invention. Note: The complete sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 29105 BP; 5477 A; 7362 C; 7988 G; 5611 T; 0 U; 2667 Other;
XX SQ
XX
XX Query Match 80.8%; Score 20.2; DB 9; Length 29105;
XX Best Local Similarity 88.0%; Pred. No. 5.7e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAATCCGACGACAAATCT 25
XX Db 28585 AAAAAAAAAAGCAGACAGATCT 28609
XX
XX RESULT 24
XX ADB72182
XX ID ADB72182 standard; DNA; 29105 BP.
XX AC
XX AD B72182;
XX DT 04-DEC-2003 (first entry)
```

```
XX XX
XX DE Human MAP3K6 gene.
XX DE
XX DE human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX DE cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX OS
XX OS Homo sapiens.
XX OS
XX OS WO2003008583-A2.
XX OS
XX OS 30-JAN-2003.
XX OS
XX OS 26-DEC-2001; 2001WO-US051291.
XX OS
XX OS 02-MAR-2001; 2001US-00798586.
XX OS
XX OS 23-OCT-2001; 2001US-00004113.
XX OS
XX OS 08-NOV-2001; 2001US-00052482.
XX OS
XX OS 30-NOV-2001; 2001US-00997722.
XX OS
XX OS 20-DEC-2001; 2001US-00034650.
XX OS
XX OS (SAGR-) SAGRES DISCOVERY.
XX OS
XX OS Morris DW; Engelhard BK;
XX OS
XX OS WPI; 2003-239337/23.
XX OS
XX PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX PT cancers, neoplasm, adenocarcinoma, or sarcoma.
XX PT
XX PS Claim 1; SEQ ID NO 10; 2304bp; English.
XX PS
XX CC The invention relates to a novel recombinant nucleic acid comprising a
XX CC nucleotide sequence selected from any of the 660 sequences fully defined
XX CC in the specification. A polynucleotide of the invention has cytostatic
XX CC activity, and may have a use in gene therapy, or in a vaccine. The
XX CC recombinant nucleic acids and polypeptides are useful for treating
XX CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX CC sarcomas. The present sequence represents a human gene of the invention.
XX CC
XX SQ Sequence 29105 BP; 5477 A; 7362 C; 7988 G; 5611 T; 0 U; 2667 Other;
XX SQ
XX
XX Query Match 80.8%; Score 20.2; DB 10; Length 29105;
XX Best Local Similarity 88.0%; Pred. No. 5.7e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAATCCGACGACAAATCT 25
XX Db 28585 AAAAAAAAAAGCAGACAGATCT 28609
XX
XX RESULT 25
XX ABV75372
XX ID ABV75372 standard; DNA; 31024 BP.
XX AC
XX AC ABV75372;
XX AC
XX DT 07-MAR-2003 (first entry)
XX DE
XX DE Human IGFBP-2 gene sequence.
XX DE
XX DE Insulin-like growth factor binding protein; IGFBP; cytostatic; liver;
XX DE cancer; human; IGFBP-2; gene; ds.
XX OS
XX OS Homo sapiens.
XX OS
XX OS Key Location/Qualifiers
XX OS 115..30695
XX OS CDS
XX FT
XX FT //tag= a
XX FT //product= "IGFBP-2"
XX FT //note= "Insulin-like growth factor binding protein;
XX FT contains introns"
XX FT 115..556
XX FT exon
XX FT //tag= b
```

```
FT FT /number= 1
FT intron 557. .27147
FT /tag= c
FT /number= 1
FT exon 27148. .27377
FT /tag= d
FT /number= 2
FT intron 27378. .28448
FT /tag= e
FT /number= 2
FT exon 28449. .28589
FT /tag= f
FT intron 28590. .30530
FT /tag= g
FT /number= 3
FT exon 30531. .30692
FT /tag= h
FT /number= 4
XX PN WO200230580-A1.
XX PD 14-NOV-2002.
XX PF 03-MAY-2002; 2002WO-AU000556.
XX PR 03-MAY-2001; 2001US-0288441P.
XX PA (NACA-) NAT CANCER CENT SINGAPORE PTE LTD.
XX PA (SICE-) SINGAPORE GEN HOSPITAL PTE LTD.
XX PA (ARGA/) ARGAE T V P.
XX PI Huynh TH, Chow PKH, Soo KC;
XX DR WPI: 2003-103522/09.
XX P-PEDB; ABB82758.
XX PT Detecting the presence or diagnosing the risk of a liver cancer in a
XX PT patient comprises detecting aberrant expression of a gene encoding an
XX PT insulin-like growth factor binding protein.
XX PS Example; Page 110-127; 142pp; English.
XX CC The invention relates to detecting the presence or diagnosing the risk of
XX CC a liver cancer in a patient. The method involves detecting in a
XX CC biological sample obtained from the patient aberrant expression of a gene
XX CC encoding an insulin-like growth factor binding protein (IGFBP). The
XX CC method is useful for detecting the presence or diagnosing the risk of a
XX CC liver cancer or for screening agents in a patient. The agent is useful
XX CC for the manufacture of a medicament for treating and/or preventing liver
XX CC cancer. The present sequence represents a human IGFBP-2 polypeptide
XX CC encoding genomic DNA (GenBank Accession No. gi/18552832:211494-242517)
XX SQ Sequence 31024 BP; 7114 A; 7325 C; 8230 G; 8355 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 10; Length 31024;
Best Local Similarity 88.0%; Pred. No. 5.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAATCGACGACAAATCT 25
Db 10259 AAAAAAAAAATTGCAGAAAAAATCT 10283
RESULT 26
ID AAK86139 standard; DNA; 37736 BP.
XX AAK86139;
AC AAK86139;
XX 07-NOV-2001 (first entry)
DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40951.
XX
```

```
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001354.
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214866P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 14-JUL-2000; 2000US-0217496P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225214P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 14-AUG-2000; 2000US-0225758P.
XX PR 14-AUG-2000; 2000US-0225759P.
XX PR 18-AUG-2000; 2000US-0226279P.
XX PR 22-AUG-2000; 2000US-0226681P.
XX PR 22-AUG-2000; 2000US-0226686P.
XX PR 23-AUG-2000; 2000US-0227182P.
XX PR 30-AUG-2000; 2000US-0228924P.
XX PR 01-SEP-2000; 2000US-0229287P.
XX PR 01-SEP-2000; 2000US-0229343P.
XX PR 01-SEP-2000; 2000US-0229344P.
XX PR 01-SEP-2000; 2000US-0229345P.
XX PR 05-SEP-2000; 2000US-0229509P.
XX PR 05-SEP-2000; 2000US-0229513P.
XX PR 06-SEP-2000; 2000US-0230437P.
XX PR 06-SEP-2000; 2000US-0230438P.
XX PR 08-SEP-2000; 2000US-0231242P.
XX PR 08-SEP-2000; 2000US-0231243P.
XX PR 08-SEP-2000; 2000US-0231244P.
XX PR 08-SEP-2000; 2000US-0231413P.
XX PR 08-SEP-2000; 2000US-0231414P.
XX PR 08-SEP-2000; 2000US-0232080P.
XX PR 08-SEP-2000; 2000US-0232081P.
XX PR 12-SEP-2000; 2000US-0231968P.
XX PR 14-SEP-2000; 2000US-0232397P.
XX PR 14-SEP-2000; 2000US-0232398P.
XX PR 14-SEP-2000; 2000US-0232399P.
XX PR 14-SEP-2000; 2000US-0232400P.
XX PR 14-SEP-2000; 2000US-0232401P.
XX PR 14-SEP-2000; 2000US-0233063P.
XX PR 14-SEP-2000; 2000US-0233064P.
XX PR 14-SEP-2000; 2000US-0233065P.
XX PR 21-SEP-2000; 2000US-0234223P.
```

PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0251989P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 DR
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PR useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS
 XX Disclosure; SEQ ID NO 40951; 3071bp + Sequence Listing; English.
 CC
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 37736 BP; 9869 A; 9250 C; 9209 G; 9408 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 4; Length 37736;
 Best Local Similarity 88.0%; Pred. No. 5,7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAATCGAGCAATCT 25
 Db 5567 AAAAAAAAAATCGCAAAATATCT 5591
 RESULT 27
 ABL67924/c
 ID ABL67924 standard; DNA; 41936 BP.
 XX
 AC ABL67924;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Ovary cancer related gene sequence SEQ ID NO:6261.
 XX
 KM Human; cancer; colon; breast; ovary; esophagus; kidney; thyroid;
 KM stomach; lung; prostate; pancreas; carcinoma; antitumour; Cancerous;
 KM cytosaratic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KM gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US010838.
 XX
 PR 05-JUN-2000; 2000US-0209473P.
 PR 18-SEP-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233133P.
 PR 20-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.

PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235180P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.

XX (AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;

XX WPI: 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
XX in expression of a gene of a signature gene set.

PS Claim 1; SEQ ID NO 6261; 44pp; English.

XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in AB161664
CC to AB170110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophagal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
CC tumour

XX Sequence 41936 BP; 9311 A; 11171 C; 11849 G; 9605 T; 0 U; 0 Other;

XX Query Match 80.8%; Score 20.2; DB 6; Length 41936;
XX Best Local Similarity 88.0%; Pred. No. 5.8e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 34468 AAAAAAAAAATCGCAGCAAAATCT 34444

RESULT 28

Continuation (12 of 17) of ABX08336 from base 1100001 (Human phosphodiesterase 4D (PDE4I
WP Sequence split into 17 fragments LOCUS ABX08336 Accession ABX08336

Fragment Name	Begin	End
WP ABX08336_00	1	110000
WP ABX08336_01	100001	210000
WP ABX08336_02	200001	310000
WP ABX08336_03	300001	410000
WP ABX08336_04	400001	510000
WP ABX08336_05	500001	610000
WP ABX08336_06	600001	710000
WP ABX08336_07	700001	810000
WP ABX08336_08	800001	910000
WP ABX08336_09	900001	1010000
WP ABX08336_10	1000001	1110000
WP ABX08336_11	1100001	1210000
WP ABX08336_12	1200001	1310000
WP ABX08336_13	1300001	1410000
WP ABX08336_14	1400001	1510000
WP ABX08336_15	1500001	1610000
WP ABX08336_16	1600001	1691080

Query Match 80.8%; Score 20.2; DB 6; Length 110000;
Best Local Similarity 88.0%; Pred. No. 6e+02; Mismatches 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 67502 AAAAAAAAAATCGCAGCAAAATTT 67478

RESULT 29

ADJ25985_11/c
Continuation (12 of 17) of ADJ25985 from base 1100001 (Human phosphodiesterase 4D (PDE4I
WP Sequence split into 17 fragments LOCUS ADJ25985 Accession Adj25985

Fragment Name	Begin	End
WP ADJ25985_00	1	110000
WP ADJ25985_01	100001	210000
WP ADJ25985_02	200001	310000
WP ADJ25985_03	300001	410000
WP ADJ25985_04	400001	510000
WP ADJ25985_05	500001	610000
WP ADJ25985_06	600001	710000
WP ADJ25985_07	700001	810000
WP ADJ25985_08	800001	910000
WP ADJ25985_09	900001	1010000
WP ADJ25985_10	1000001	1110000
WP ADJ25985_11	1100001	1210000
WP ADJ25985_12	1200001	1310000
WP ADJ25985_13	1300001	1410000
WP ADJ25985_14	1400001	1510000
WP ADJ25985_15	1500001	1610000
WP ADJ25985_16	1600001	1691139

Query Match 80.8%; Score 20.2; DB 12; Length 110000;
Best Local Similarity 88.0%; Pred. No. 6e+02; Mismatches 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 67561 AAAAAAAAAATCGCAGCAAAATTT 67537

RESULT 30

ADN97989_11/c
Continuation (12 of 17) of ADN97989 from base 1100001 (Human phosphodiesterase 4D (PDE4I
WP Sequence split into 17 fragments LOCUS ADN97989 Accession Adn97989

WP Fragment Name Begin End
WP ADN97989_00 1 110000
WP ADN97989_01 100001 210000
WP ADN97989_02 200001 310000
WP ADN97989_03 300001 410000
WP ADN97989_04 400001 510000
WP ADN97989_05 500001 610000
WP ADN97989_06 600001 710000
WP ADN97989_07 700001 810000
WP ADN97989_08 800001 910000
WP ADN97989_09 900001 1010000
WP ADN97989_10 1000001 1110000
WP ADN97989_11 1100001 1210000
WP ADN97989_12 1200001 1310000
WP ADN97989_13 1300001 1410000
WP ADN97989_14 1400001 1510000
WP ADN97989_15 1500001 1610000
WP ADN97989_16 1600001 1691138

Query Match 80.8%; Score 20.2; DB 12; Length 110000;
Best Local Similarity 88.0%; Pred. No. 6e+02; 3; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGAGCAAAATCT 25
DB 67560 AAAAAAAAAATCGAGCAAAATTT 67536

RESULT 31
AD050281_11/c
Continuation (12 of 17) of AD050281 from base 1100001 (Human phosphodiesterase 4D (PDE4D
WP Sequence split into 17 fragments LOCUS AD050281 Accession AD050281

WP Fragment Name Begin End
WP AD050281_00 1 110000
WP AD050281_01 100001 210000
WP AD050281_02 200001 310000
WP AD050281_03 300001 410000
WP AD050281_04 400001 510000
WP AD050281_05 500001 610000
WP AD050281_06 600001 710000
WP AD050281_07 700001 810000
WP AD050281_08 800001 910000
WP AD050281_09 900001 1010000
WP AD050281_10 1000001 1110000
WP AD050281_11 1100001 1210000
WP AD050281_12 1200001 1310000
WP AD050281_13 1300001 1410000
WP AD050281_14 1400001 1510000
WP AD050281_15 1500001 1610000
WP AD050281_16 1600001 1691134

Query Match 80.8%; Score 20.2; DB 12; Length 110000;
Best Local Similarity 88.0%; Pred. No. 6e+02; 3; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGAGCAAAATCT 25
DB 67559 AAAAAAAAAATCGAGCAAAATTT 67535

RESULT 32
AEB85185_11/c
Continuation (12 of 17) of AEB85185 from base 1100001 (Human phosphodiesterase 4D gene S
WP Sequence split into 17 fragments LOCUS AEB85185 Accession Aeb85185

WP Fragment Name Begin End
WP AEB85185_00 1 110000
WP AEB85185_01 100001 210000
WP AEB85185_02 200001 310000
WP AEB85185_03 300001 410000
WP AEB85185_04 400001 510000
WP AEB85185_05 500001 610000
WP AEB85185_06 600001 710000
WP AEB85185_07 700001 810000

WP AEB85185_08 800001 910000
WP AEB85185_09 900001 1010000
WP AEB85185_10 1000001 1110000
WP AEB85185_11 1100001 1210000
WP AEB85185_12 1200001 1310000
WP AEB85185_13 1300001 1410000
WP AEB85185_14 1400001 1510000
WP AEB85185_15 1500001 1610000
WP AEB85185_16 1600001 1691140

Query Match 80.8%; Score 20.2; DB 14; Length 110000;
Best Local Similarity 88.0%; Pred. No. 6e+02; 3; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGAGCAAAATCT 25
DB 67562 AAAAAAAAAATCGAGCAAAATTT 67538

RESULT 33
ABD33075/c
ID ABD33075 standard; DNA; 152759 BP.

XX ABD33075;

XX 18-NOV-2004 (first entry)

XX Human cancer-associated (CA) gene HD07-002.

XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;

XX ds; cancer; cytostatic.

XX Homo sapiens.

XX WO2004058146-A2.

XX 15-JUL-2004.

XX 15-DEC-2003; 2003WO-US040081.

XX 17-DEC-2002; 2002US-00322281.

XX (SAGR-) SAGRES DISCOVERY INC.

XX Morris DW, Malandro MS;

XX WPI; 2004-499109/47.

XX Novel human cancer associated protein encoded within open reading frame

XX of cancer associated gene, useful as targets for diagnosing cancer.

XX Claim 16; SEQ ID NO 10; 182pp; English.

XX The invention relates to cancer-associated proteins (CAP) and the cancer-
XX associated (CA) nucleic acids encoding them. The invention also relates
XX to a method for treating cancers involving administering to a patient an
XX inhibitor of CAP, and a method of screening for anticancer activity in a
XX potential drug involving providing a cell that expresses a CA gene,
XX contacting a tissue sample derived from a cancer cell with an anticancer
XX drug candidate and monitoring the effect of the anticancer drug candidate
XX on expression of the CA gene. The CAP proteins are useful for detecting
XX cancer associated with expression of a CAP protein in a test cell sample
XX and for screening for a bioactive agent capable of modulating the
XX activity of a CAP protein. The CA nucleic acids are useful for diagnosing
XX cancer, involving determining the expression of a CA nucleic acid in a
XX tissue. This sequence represents a human CA gene of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 152759 BP; 43847 A; 29952 C; 31446 G; 47160 T; 0 U; 354 Other;

XX Query Match 80.8%; Score 20.2; DB 13; Length 152759;

Best Local Similarity 88.0%; Pred. No. 6.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGACAAATCT 25
DB 32201 AAAAAAAAAATCGACAAATTT 32177

RESULT 34
ABD33268/c
ID ABD33268 standard; DNA; 180227 BP.

AC ABD33268;

DT 18-NOV-2004 (first entry)

DE Human cancer-associated (CA) gene HD07-046.

DE Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;

KW ds; cancer; cytostatic.

OS Homo sapiens.

PN WO2004058146-A2.

PD 15-JUL-2004.

PF 15-DEC-2003; 2003WO-US040081.

PR 17-DEC-2002; 2002US-00322281.

PA (SAGR-) SAGRES DISCOVERY INC.

PI Morris DW, Malandro MS;

DR WPI; 2004-499109/47.

PT Novel human cancer associated protein encoded within open reading frame
of cancer associated gene, useful as targets for diagnosing cancer.

PS Claim 16; SEQ ID NO 308; 182bp; English.

CC The invention relates to cancer-associated proteins (CAP) and the cancer-
associated (CA) nucleic acids encoding them. The invention also relates
to a method for treating cancers involving administering to a patient an
inhibitor of CAP, and a method of screening for anticancer activity in a
potential drug involving providing a cell that expresses a CA gene,
contracting a tissue sample derived from a cancer cell with an anticancer
drug candidate and monitoring the effect of the anticancer drug candidate
on expression of the CA gene. The CAP proteins are useful for detecting
cancer associated with expression of a CAP protein in a test cell sample
and for screening for a bioactive agent capable of modulating the
activity of a CAP protein. The CA nucleic acids are useful for diagnosing
cancer, involving determining the expression of a CA nucleic acid in a
tissue. This sequence represents a human CA gene of the invention. Note:
The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences

CC Sequence 180227 BP; 51503 A; 33645 C; 36852 G; 55474 T; 0 U; 2753 Other;

Query Match 80.8%; Score 20.2; DB 13; Length 180227;

Best Local Similarity 88.0%; Pred. No. 6.1e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGACAAATCT 25

DB 152244 AAAAAAAAAATCGACAAATCT 152220

RESULT 35

ADS31291

ID ADS31291 standard; DNA; 230 BP.

XX

AC ADS31291;

XX

DT 18-NOV-2004 (first entry)

XX

DE Human genome high complexity repeat found in the HIRA gene #324.

XX

KW Human;

XX

KW histone cell cycle regulation defective; S. cerevisiae homologue A; HIRA;

XX

KW high complexity repeat; in situ hybridisation; Southern blot;

XX

KW chromosome breakpoint; inherited genetic disease; neoplastic disorder;

XX

KW chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.

OS

Homo sapiens.

PN

US200324356-A1.

PD

04-DEC-2003.

PF

14-MAY-2001; 2001US-00854867.

PR

16-MAY-2000; 2000US-00573080.

PA

(KNOL/) KNOL J H M.

PI

(ROGA/) ROGAN P K.

DR

WPI; 2002-062378/08.

PT

Single copy genomic hybridization probes for detecting specific nucleic
acid sequences in sample by in situ hybridization useful for detection of
acquired or inherited genetic diseases.

PS

Example 1; SEQ ID NO 324; 30bp; English.

CC The invention relates to a nucleic acid hybridisation probe comprising a
CC labelled, single copy nucleic acid of at least 50 nucleotides, which
CC will hybridise to a deduced single copy sequence interval in target
CC nucleic acid (TNA) of known sequence. The single copy sequence is deduced
CC by comparing the target nucleic acid (e.g. a disease causing gene) with a
CC collection of high and low complexity repeat sequences as found in the
CC genome of the organism from containing the target nucleic acid. The probe
CC is generated by PCR on the target sequence. The probe is essentially free
CC of blocking nucleic acid sequences which will hybridise to repeat
CC sequences within the genome of which the TNA is a part, and is labelled
CC with a label selected from fluorochrome-responsive labels, fluorochromes,
CC calorimetric chemical, conjugated proteins, antibodies, antigens and
CC their mixtures. The probe is useful in a hybridisation method, where the
CC hybridisation method is from in situ hybridisation, Southern blot, and
CC other methods in which nucleic acid is immobilised, where the method
CC further comprises selecting a single copy nucleic acid which will
CC hybridise to a duplilon or triplicon sequence domain. The probe is useful
CC for determining the existence of previously unknown repeat sequence
CC families in a genome. The method comprises reacting a labelled probe with
CC the genome, causing the probe to hybridise and ascertaining if the probe
CC hybridises to the genome at more than three preferably ten different
CC locations as a determination of new repeat sequence family, where the
CC determining step comprises selecting the single copy sequence from a
CC duplilon or triplicon sequence domain. The probe is useful for
CC determining a chromosome breakpoint and is useful in the fields for
CC cytogenetics and molecular genetics for determining the presence of
CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
CC the probes may be used to analyse specific chromosomal locations by in
CC situ hybridisation as a detection of acquired or inherited genetic
CC diseases especially for detection of genetic or neoplastic disorders.
CC Unlike prior art techniques, the probe permits more precise chromosomal
CC breakpoint determinations by in situ hybridisation. The genomic sequence
CC comprising the human HIRA gene (histone cell cycle regulation defective,
CC S. cerevisiae, homologue A) was analysed for single copy sequence
CC intervals for use as probes of the invention. HIRA is located on
CC chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and
CC Velo-Cardio-facial syndromes. The present sequence is a high complexity

CC repeat found within the human genome used to analyse the HIRA gene for
 CC repeat regions. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030224356.
 XX

SQ Sequence 230 BP; 70 A; 51 C; 55 G; 51 T; 0 U; 3 Other;

Query Match 79.2%; Score 19.8; DB 7; Length 230;
 Best Local Similarity 84.0%; Pred. No. 6.4e+02;
 Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
 |||||:|||||
 DB 109 AAAAAAAAAATCCGMAAAATCT 133

RESULT 36

ADY36679
 ID ADY36679 standard; DNA; 230 BP.

AC ADY36679;

DT 05-MAY-2005 (first entry)

DE HIRA genomic fragment SEQ ID NO 324.

XX hybridization; DNA detection; neoplasm; genetic disorder; cytogenetics;
 XX HIRA; ds.

OS Homo sapiens.

PN WO20018089-A2.

PD 22-NOV-2001.

PE 15-MAY-2001; 2001WO-US015674.

PF 16-MAY-2000; 2000US-00573080.

PR 14-MAY-2001; 2001US-00854867.

XX (CHIL-) CHILDREN'S MERCY HOSPITAL.

XX Knoi1 JHM, Rogan PK, Cazarro PM;

DR WPI; 2002-062378/08.

PT Single copy genomic hybridization probes for detecting specific nucleic
 PT acid sequences in sample by in situ hybridization useful for detection of
 PT acquired or inherited genetic diseases.

XX Example 1; SEQ ID NO 324; 67pp; English.

CC The invention describes a nucleic acid hybridization probe (I) comprising
 CC a labeled, single copy nucleic acid of at least 50 nucleotides, which
 CC will hybridize to a deduced single copy sequence interval in target
 CC nucleic acid (TNA) of known sequence. (I) is useful in a hybridization
 CC method which comprises preparing a reaction mixture comprising TNA and
 CC (I) which hybridizes to TNA, and causing (I) to hybridize to TNA, where
 CC the hybridization method is from in situ hybridization, Southern blot,
 CC and other methods in which nucleic acid is immobilized, where the method
 CC further comprises selecting a single copy nucleic acid which will
 CC hybridize to a dupliron or triplicon sequence domain. (I) is useful for:
 CC determining the existence of previously unknown repeat sequence families
 CC in a genome; determining a chromosome breakpoint and in the fields of
 CC cytogenetics and molecular genetics for determining the presence of
 CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
 CC the probes may be used to analyze specific chromosomal locations by in
 CC situ hybridization as a detection of acquired or inherited genetic
 CC diseases especially for detection of genetic or neoplastic disorders.
 CC Unlike prior art techniques, (I) permits more precise chromosomal
 CC breakpoint determinations by in situ hybridization. Hybridization
 CC techniques utilizing (I), have made it possible to obtain reliable,
 CC easily detectable signals with relatively small probes. A readily

CC detectable signal was obtained with a probe on the order of 2 kb in
 CC length, using fluorescent in situ hybridization (FISH) technology. This
 CC sensitivity of (I) is improved compared to the prior art, because the
 CC probes of (I) are homogeneous single copy sequences. However, smaller
 CC amplified segments, each comprising non-repetitive sequences, may also be
 CC used in combination as probes to achieve adequate signals for in situ
 CC hybridization. Complex single copy probes that hybridize to duplicated or
 CC triplicated targets can also increase hybridization signals. This
 CC sequence represents a human HIRA genomic sequence that shows homology to
 CC a known high-complexity repeat sequence family of the human genome and is
 CC used in the creation of an HIRA gene probe.
 XX

SQ Sequence 230 BP; 70 A; 51 C; 55 G; 51 T; 0 U; 3 Other;

Query Match 79.2%; Score 19.8; DB 7; Length 230;
 Best Local Similarity 84.0%; Pred. No. 6.4e+02;
 Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
 |||||:|||||
 DB 109 AAAAAAAAAATCCGMAAAATCT 133

RESULT 37

ACNS1137
 ID ACNS1137 standard; cDNA; 267 BP.

AC ACNS1137;

DT 02-DEC-2004 (first entry)

DE Cotton androecium tissue EST Clone ID: L183828-006-Q1-K6-B2, SEQ:5918.

XX

KM Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
 KM variety Nuccotton33B; library L183828; molecular tag; molecular marker;

KW genetic mapping; molecular mapping; seed germination; plant growth;
 KW plant quality; plant yield; plant breeding; tissue printing; ss.

XX Gossgypium hirsutum.

XX US2004123340-A1.

PD 24-JUN-2004.

PE 12-DEC-2001; 2001US-00021323.

PF 14-DEC-2000; 2000US-0255619P.

XX (DEIK/) DEIKMAN J.

PA (FENG/) FENG P C C.

PA (FINCH/) FINCHER K L.

PA (ZIEG/) ZIEGLER T E.

XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;

XX WPI; 2004-479808/45.

XX Claim 1; SEQ ID NO 5918; 34pp; English.

CC The invention relates to 17880 cotton expressed sequence tags (ESTs);
 CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
 CC from primed or non-primed seeds from variety DP50B, mature seeds from
 CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
 CC tissue, developing fibres, carpel walls and septa from variety
 CC Nuccotton33B. The invention also relates to substantially purified
 CC nucleotons or their fragments encoded by nucleic acid molecules of the
 CC invention, and to transformed plants having a nucleic acid construct
 CC comprising a nucleic acid of the invention. The cotton ESTs are useful as

CC	molecular tags to isolate genetic regions, to isolate genes, to map
CC	genes, to determine gene function and to determining whether genes are
CC	members of a particular gene family. The nucleic acid molecules may be
CC	used for isolating a variety of agronomically significant genes
CC	associated with plant growth, quality, yield, and could also serve as
CC	links in metabolic and catabolic pathways. The nucleic acid molecules are
CC	also useful for identifying genes important in initiating and maintaining
CC	seed germination or that may be used to mitigate stresses encountered
CC	during seed germination. The ESTs additionally enable the acquisition of
CC	promoters and cis-regulatory elements which will be useful to express
CC	agronomically significant genes in these tissues and/or other tissues,
CC	and also permits the acquisition of molecular markers useful in breeding
CC	schemes, genetic and molecular mapping, and in cloning of agronomically
CC	significant genes. The nucleic acid molecules are further useful for
CC	detecting the expression level or pattern of a protein or mRNA and for
CC	detecting the presence or quantity of a protein by tissue printing. The
CC	present sequence represents a specifically claimed EST isolated from a
CC	cotton variety Nucleon33B androecium tissue cDNA library (LIB3628). The
CC	sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the US
CC	patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340
CC	
SO	Sequence 267 BP; 135 A; 32 C; 42 G; 56 T; 0 U; 2 Other;
	Query Match 79.2%; Score 19.8; DB 13; Length 267;
	Best Local Similarity 91.3%; Pred. No. 6.4e+02;
	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0
Qy	1 AAAAAAAAAATCGGACCAAT 23
Db	199 AAAAAAAAAAGCAGCAAAAT 221
RESULT 38	
ABQ21911	
ID	ABQ21911 standard; DNA; 577 BP.
XX	
AC	ABQ21911;
XX	
DT	12-JUL-2002 (first entry)
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 8502.
XX	
KW	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW	drug; side effect; cancer; central nervous system; cardiovascular;
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;
KW	SNP; cell differentiation; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200218632-A2.
XX	
PD	07-MAR-2002.
XX	
PF	01-SEP-2001; 2001WO-EP010074.
XX	
PR	01-SEP-2000; 2000DE-01043826.
PR	05-SEP-2000; 2000DE-01044543.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Pleepenbrock C, Berlin K, Gueutig D;
XX	
DR	WPI, 2002-371829/40.
XX	
PT	Determining the degree of cytosine methylation in genomic DNA, useful for
PT	diagnosis and prognosis, comprises selective hybridization of amplicons
XX	from chemically treated DNA.
XX	
BS	Claim 12; 56pp + Sequence Listing; 56pp; German.
XX	
CC	This invention describes a novel method for determining the degree of
CC	methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC	genomic sample of DNA. The sample is treated chemically to convert
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
CC	DNA that contains the target C is amplified to form a labeled amplicon.
CC	The amplicon is hybridised to two classes, each with at least one member,
CC	of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC	degree of hybridisation to both classes is determined from the label on
CC	the amplicon. From the ratio of labels hybridised to the two classes of
CC	oligomers, the degree of methylation is calculated. The method is used:
CC	(1) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC	and of a wide range of diseases, e.g. cancer, disorders of the central
CC	nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC	particularly by detecting mutations or single nucleotide polymorphisms
CC	(SNP's); and (ii) for differentiation of cell or tissue types and for
CC	investigating cell differentiation. The method allows the methylation
CC	status of many C residues to be determined simultaneously. ABO13410-
CC	ABO54121 represent genomic DNA sequences used to illustrate the method
CC	for determining the degree of cytosine methylation described in the
CC	disclosure of the invention
SQ	Sequence 577 BP; 306 A; 78 C; 45 G; 148 T; 0 U; 0 Other;
Query Match	79.2%; Score 19.8; DB 6; Length 577;
Best Local Similarity	91.3%; Pred. No. 6.7e+02;
Matches	21; Conservative 0; Mismatches 2; Indels 0; Gaps 0.
QY	1 AAAAAAAAAATCCAGACAAAT 23
DB	380 AAAAAAAAAATCGACGCAAAAT 402
RESULT 39	
ABO21910/C	
ID	ABO21910 standard; DNA; 577 BP.
AC	ABO21910;
XX	
DT	12-JUL-2002 (first entry)
XX	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 8501.
XX	
KM	Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;
KW	drug; side effect; cancer; central nervous system; cardiovascular;
KW	gastrointestinal; respiratory system; single nucleotide polymorphism;
XX	SNP; cell differentiation; de.
OS	Homo sapiens.
PN	WO200218632-A2.
PD	07-MAR-2002.
PX	
PF	01-SEP-2001; 2001WO-EP010074.
XX	
PR	01-SEP-2000; 2000DE-01043826.
PR	05-SEP-2000; 2000DE-01044543.
XX	
PA	(EPIG-) EPIGENOMICS AG.
PI	Olek A. Piepenbrock C, Berlin K, Guetig D;
DR	WPI; 2002-371829/40.
PT	
XX	
PT	Determining the degree of cytosine methylation in genomic DNA, useful for
PT	diagnosis and prognosis, comprises selective hybridization of amplicons
PT	from chemically treated DNA.
PS	Claim 12; 56pp + Sequence Listing; 56pp; German.
XX	
CC	This invention describes a novel method for determining the degree of
CC	methylation of a particular cytosine in a motif 5'-CPG-3', present in a
CC	genomic sample of DNA. The sample is treated chemically to convert
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
CC	DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's), and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. AB013410-
CC AB054121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention

XX Sequence 577 BP; 148 A; 45 C; 78 G; 306 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 577;
Best Local Similarity 91.3%; Pred. No. 6.7e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAAATCGACGACCAAT 23

198 AAAAAAAAAATCGACGACCAAT 176

RESULT 40

ABL34385/C
ID ABL34385 standard; DNA; 1710 BP.

XX ABL34385;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 2358.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianemic; cytosine; noctropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antineumatic; antirheumatic; antidiabetic; antipneumonia;
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIDEMIOLOGICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX PT diagnosis and treatment of diseases associated with abnormal cytosine
XX PT methylation.

XX Claim 1; SEQ ID NO 2358; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX

XX Sequence 1710 BP; 378 A; 71 C; 433 G; 828 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 1710;
Best Local Similarity 91.3%; Pred. No. 7e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAAATCGACGACCAAT 23

469 AAAAAAAAAATCGACGACCAAT 447

RESULT 41

AAV83941/C
ID AAV83941 standard; DNA; 5596 BP.

XX AAV83941;

XX 03-MAR-1999 (first entry)

XX Bacterial artificial chromosome (BAC)-F2 contig 1.

XX Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
XX neocentromere; replication; extra-chromosomal element; segregation;
XX cell division; artificial chromosome; gene therapy; BAC; transgenic;
XX human artificial chromosome; bacterial artificial chromosome; ss.

XX Synthetic.

XX WO9851790-A1.

XX 19-NOV-1998.

XX 13-MAY-1998; 98NO-AU000352.

XX 13-MAY-1997; 97AU-00006784.

XX 26-AUG-1997; 97AU-00008791.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Choo K, Du Sart D, Cancilla MR;

XX WPI; 1999-009773/01.

XX New isolated nucleic acid comprising neocentromere sequences from
XX PT eukaryotic chromosome - used to produce replicable, segregating
XX PT artificial chromosomes that can carry large amounts of DNA for gene
XX PT therapy.

XX Claim 10; Page 177-181; 540pp; English.

XX The present sequence represents a bacterial artificial chromosome (BAC)
XX CC contig, and exemplifies the invention. The specific location describes
XX CC nucleic acid sequences derived from a eukaryotic chromosome, including a
XX CC neocentromere or its functional derivative or hybrid, that are able, in a
XX CC compatible cell, of replicating, acting as extra-chromosomal element and
XX CC segregating during cell division. The sequences can be used to construct
XX CC artificial chromosomes for use in gene therapy comprising a replicable,
XX CC segregating nucleic acid that confers a specific phenotype on cells.
XX CC Human artificial chromosomes can propagate in human cells and carry large
XX CC amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
XX CC they are not mutagenic. The artificial chromosomes are also useful for
XX CC generation of transgenic plants and animals, in production of proteins
XX CC and to make diagnostic reagents, e.g. for expression of cytokines,
XX CC receptors and growth factors, or to increase the copy number of a gene in
XX CC a cell. The constructs may also be used for functional and structural
XX CC analysis of chromosomes

XX Sequence 5596 BP; 1730 A; 1019 C; 1082 G; 1763 T; 0 U; 2 Other;

```
Query Match          79.2%; Score 19.8; DB 2; Length 5596;
Best Local Similarity 91.3%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAATCGACAAATC 24
   |||||
Db 711 AAAAAAAAAATCAAGACAAATC 689

RESULT 42
ADQ97173/C
ID ADQ97173 standard; DNA; 40783 BP.
XX
AC ADQ97173;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD08-010, SEQ ID 149.
XX
KM Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 149; 1999p; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 40783 BP; 11403 A; 8254 C; 8101 G; 13025 T; 0 U; 0 Other;

Query Match          79.2%; Score 19.8; DB 12; Length 40783;
Best Local Similarity 91.3%; Pred. No. 8.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACAAAT 23
   |||||
Db 30361 AAAAAAAAAATCTCAAAAT 30339

RESULT 43
AAV83940/C
ID AAV83940 standard; DNA; 80240 BP.
XX
AC AAV83940;
XX
DT 03-MAR-1999 (first entry)
XX
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KM Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
```

```
KM neocentromere; replication; extra-chromosomal element; segregation;
KM cell division; artificial chromosome; gene therapy; mardel(10);
KM human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
OS Homo sapiens.
PN WO9851790-A1.
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
XX
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
XX
XX Claim 9; Fig 16A; 540pp; English.
XX
CC The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC mitotically stable and contains a functional neocentromere at a location
CC regarded as non-centromeric. This neocentromere maps to q25.2 on
CC chromosome 10. The specification describes nucleic acid sequences derived
CC from a eukaryotic chromosome, including a neocentromere or its functional
CC derivative or hybrid, that are able, in a compatible cell, of
CC replicating, acting as extra-chromosomal element and segregating during
CC cell division. The sequences can be used to construct artificial
CC chromosomes for use in gene therapy comprising a replicable, segregating
CC nucleic acid that confers a specific phenotype on cells. Human artificial
CC chromosomes can propagate in human cells and carry large amounts of DNA
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC mutagenic. The artificial chromosomes are also useful for generation of
CC transgenic plants and animals, in production of proteins and to make
CC diagnostic reagents, e.g. for expression of cytokines, receptors and
CC growth factors, or to increase the copy number of a gene in a cell. The
CC constructs may also be used for functional and structural analysis of
CC chromosomes
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

Query Match          79.2%; Score 19.8; DB 2; Length 80240;
Best Local Similarity 91.3%; Pred. No. 8.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAATCGACAAATC 24
   |||||
Db 68701 AAAAAAAAAATCAAGACAAATC 68679

RESULT 44
AAV83939/C
ID AAV83939 standard; DNA; 80595 BP.
XX
AC AAV83939;
XX
DT 03-MAR-1999 (first entry)
XX
DE HC-contig derived from normal human chromosome 10q25.2 region.
XX
KM Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KM neocentromere; replication; extra-chromosomal element; segregation;
KM cell division; artificial chromosome; gene therapy; mardel(10);
```

human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.

XX Homo sapiens.

XX MO9851790-A1.

XX 19-NOV-1998.

XX 13-MAY-1998; 98WO-AU000352.

XX 13-MAY-1997; 97AU-00006784.

XX 26-AUG-1997; 97AU-00008791.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Choo K, Du Sart D, Cancellia MR;

XX WPI; 1999-009773/01.

XX New isolated nucleic acid comprising neocentromere sequences from

XX eukaryotic chromosomes - used to produce replicable, segregating

XX artificial chromosomes that can carry large amounts of DNA for gene

XX therapy.

XX Claim 8; Fig 6; 540pp; English.

XX The present sequence represents the HC-contig derived from normal human

XX chromosome 10, 10q25.2 region. This region can be naturally mutated to

XX produce an unusual chromosomal marker referred to as mardel1(10). The

XX mardel1(10) marker is mitotically stable and contains a functional

XX neocentromere at a location regarded as non-centromeric. This

XX neocentromere maps to q25.2 on chromosome 10. The specification describes

XX nucleic acid sequences derived from a eukaryotic chromosome, including a

XX neocentromere or its functional derivative or hybrid, that are able, in a

XX compatible cell, of replicating, acting as extra-chromosomal element and

XX segregating during cell division. The sequences can be used to construct

XX artificial chromosomes for use in gene therapy comprising a replicable,

XX segregating nucleic acid that confers a specific phenotype on cells.

XX Human artificial chromosomes can propagate in human cells and carry large

XX amounts of DNA (e.g. therapeutic genes) and, being extra-chromosomal,

XX they are not mutagenic. The artificial chromosomes are also useful for

XX generation of transgenic plants and animals, in production of proteins

XX and to make diagnostic reagents, e.g. for expression of cytokines,

XX receptors and growth factors, or to increase the copy number of a gene in

XX a cell. The constructs may also be used for functional and structural

XX analyses of chromosomes

XX

XX

XX Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

XX

XX Query Match 79.2%; Score 19.8; DB 2; Length 80595;

XX Best Local Similarity 91.3%; Pred. No. 8.3e+02;

XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX

XX 2 AAAAAAAAAATCGACGACCAATC 24

XX ||||||||| |||||||

XX Db 68964 AAAAAAAAAATCAAGACCAATC 68942

XX

XX RESULT 45

XX ADG70184_2

XX Continuation (3 of 4) of ADG70184 from base 200001 (DNA of BAC DA236mt15-00303.)

XX WP Sequence split into 4 fragments LOCUS ADG70184 Accession Adg70184

XX Fragment Name Begin End

XX WP ADG70184_0 1 110000

XX WP ADG70184_1 10001 210000

XX WP ADG70184_2 20001 310000

XX WP ADG70184_3 30001 379652

XX

XX Query Match 79.2%; Score 19.8; DB 10; Length 110000;

XX Best Local Similarity 91.3%; Pred. No. 8.4e+02;

XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX 1 AAAAAAAAAATCGACGACCAATC 23

```

Db          40885 AAAAAAAAAAATTGCAGCAACT 40907
|||||||
RESULT 46
ABN92804
ID ABN92804 standard; DNA; 498 BP.
XX
AC ABN92804;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2267.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
antibacterial; gene therapy; gene; ds.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-00134001.
XX
PR 14-AUG-1997; 97US-0055779P.
XX
PP 08-NOV-1997; 97US-0064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucet-Stamm LA, Bush D;
XX
DR MPI; 2002-381255/41.
XX
DR P-PDB; ABP40259.
PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
polypeptide, useful for diagnosing and treating bacterial infections.
PS Disclosure; SEQ ID NO 2267; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
frame (ORF) nucleic acid sequences which encode the amino acid sequences
given in ABP5124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
screen for compounds able to interfere with the S. epidermidis life cycle
or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
in electronic format directly from the USPTO web site
CC
SQ Sequence 498 BP; 195 A; 88 C; 52 G; 163 T; 0 U; 0 Other;

Query Match      76.8%; Score 19.2; DB 6; Length 498;
Best Local Similarity 87.5%; Pred.No.1,1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy   2 AAAAAAAAAATCGCAGCAAACT 25
    |||||
Db   435 AACAAAAAATGCGAGAAATCT 458

RESULT 47
ADS02942
ID ADS02942 standard; DNA; 498 BP.
XX
AC ADS02942;
XX
DT 04-NOV-2004 (first entry)
XX
DE Staphylococcus epidermis polynucleotide seqid 2237.
XX
KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
```


KW recombinant expression vector; infection; computer readable medium;
KM computer based system; gene; ds.
XX Staphylococcus epidermidis.
XX US2004147734-A1.
XX 29-JUL-2004.
XX 01-DEC-2003; 2003US-00724972.
XX 08-NOV-1997; 97US-0064964P.
XX 13-AUG-1998; 98US-00134001.
XX 29-NOV-1999; 99US-00450969.
XX (DOUC/) DOUCETTE-STAMM L.
XX (BUSH/) BUSH D.
XX Doucette-Stamm L, Bush D;
XX WPI; 2004-580138/56.
XX P-PSDB; ADS06714.
XX
XX New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.
XX
XX Claim 5; SEQ ID NO 2237; 741pp; English.
XX
XX The invention describes an isolated nucleic acid comprising a nucleotide
XX sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
XX 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
XX of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
XX given in the specification. Also described are: a recombinant expression
XX vector; a cell comprising a recombinant expression vector of (1);
XX comprising an S. epidermidis polypeptide; an isolated nucleic acid
XX comprising a nucleotide sequence of at least 8 nucleotides in length; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection, comprising a nucleic acid cited above and a carrier; treating
XX a subject for S. epidermidis infection; a recombinant or substantially
XX pure preparation of an S. epidermidis polypeptide or its fragment; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection; detecting the presence of a Staphylococcus nucleic acid in a
XX sample; a computer readable medium having recorded in it the nucleotide
XX sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
XX system for identifying fragments of the Staphylococcus genome of
XX commercial importance; a computer based system for identifying fragments
XX of the Staphylococcus plasmids of commercial importance; identifying
XX commercially important nucleic acid fragments of the Staphylococcus
XX genome and/or plasmids; and identifying an expression modulating fragment
XX of the Staphylococcus genome and/or plasmids. The methods and
XX compositions of the present invention are useful for the diagnosis,
XX prevention and/or treatment of an Staphylococcal epidermidis bacterial
XX infection. This sequence encodes a S. epidermidis protein of the invention.
XX
XX Sequence 498 BP; 195 A; 88 C; 52 G; 163 T; 0 U; 0 Other;
SQ

Query Match 76.8%; Score 19.2; DB 13; Length 498;
Best Local Similarity 87.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAATCGACGAATCT 25
DB 435 AAAAAAAAAATCGACGAATCT 458

RESULT 48
ABQ25827 standard; DNA; 544 BP.
XX
XX ABQ25827;
XX
XX 12-JUL-2002 (first entry)
XX

XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 12418.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX 05-SEP-2000; 2000DE-01044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention
SQ

Query Match 76.8%; Score 19.2; DB 6; Length 544;
Best Local Similarity 87.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGAATC 24
DB 133 AAAAAAAAAATCGACGAATC 156

RESULT 49
ABQ25826/c
ID ABQ25826 standard; DNA; 544 BP.
XX
XX ABQ25826;
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 12417.
XX

```

KW SNP: cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX
XX
XX 07-MAR-2002.
XX
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX 05-SEP-2000; 2000DE-01044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX
XX Olek A, Piepenbrock C, Berlin K, Gueig D;
XX
XX WPI: 2002-371829/40.
XX
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention
XX
XX
XX Sequence 612 BP; 67 A; 57 C; 175 G; 310 T; 0 U; 3 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 6; Length 612;
XX Best Local Similarity 87.5%; Pred. No. 1.1e+03;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0
XX
XX
XX 1 AAAAAAAAAATCGCAGCAATC 24
XX ||||||||| |||||
XX
XX Db 607 AAAAAAAAAACCGCACACAAAAC 584

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:34:03 ; Search time 1752.1 Seconds
(without alignments)
667.586 Million cell updates/sec

Title: US-10-681-773-2

Perfect score: 25

Sequence: 1 aaaaaaaaaatcgcagacaatc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gsa1:*
10: gb_gsa2:*
11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	21.8	87.2	424	9	AQ023140 HS_2177_A
C 2	21.8	87.2	459	3	BJ929968
C 3	21.8	87.2	492	3	B35642
C 4	21.8	87.2	495	9	A2068572
C 5	21.8	87.2	507	9	AQ621055
C 6	21.8	87.2	574	9	A2476425
C 7	21.8	87.2	668	9	BH512725
C 8	21.8	87.2	682	10	AG099199
C 9	21.8	87.2	795	9	BH248096
C 10	21.8	87.2	898	5	BU509223
C 11	21.4	85.6	511	10	CE581913
C 12	21.4	85.6	614	10	CE372285
C 13	21.4	85.6	706	10	CE781479
C 14	20.8	83.2	199	3	BJ683132
C 15	20.8	83.2	382	6	CD290160
C 16	20.8	83.2	477	5	BO632287
C 17	20.8	83.2	599	10	CL761036
C 18	20.8	83.2	605	7	CN006484
C 19	20.8	83.2	633	7	CV709331
C 20	20.8	83.2	675	8	CX634942
C 21	20.8	83.2	720	7	CN980172
C 22	20.8	83.2	729	8	CX670454

C	96	20.2	80.8	488	9	BZ480903	BZ480903 BOMAS888R
C	97	20.2	80.8	490	11	DS091485	DS091485 Oryzias 1
C	98	20.2	80.8	492	5	BO632089	BO632089 1122C08.r
C	99	20.2	80.8	492	8	W65367	W65367 zds2c04.r
C	100	20.2	80.8	499	9	BH038284	BH038284 RPCT-24-3
C	101	20.2	80.8	499	9	BH184449	BH184449 CH230-119
C	102	20.2	80.8	504	2	BF449964	BF449964 maad4c08.
C	103	20.2	80.8	504	7	CK143466	CK143466 SB020392
C	104	20.2	80.8	511	3	BM088127	BM088127 501517 MA
C	105	20.2	80.8	527	7	CO748106	CO748106 SNESTBaa6
C	106	20.2	80.8	532	8	DN496696	DN496696 Q007F05.5
C	107	20.2	80.8	534	5	BK640274	BK640274 BX640274
C	108	20.2	80.8	538	5	AO523511	AO523511 HS 5197 A
C	109	20.2	80.8	548	7	CV670569	CV670569 LCFE04E0
C	110	20.2	80.8	554	9	CE554181	CE554181 t1gr-g88-
C	111	20.2	80.8	559	9	AQ788186	AQ788186 HS 3135 A
C	112	20.2	80.8	578	9	AZ643057	AZ643057 IM0506C19
C	113	20.2	80.8	578	9	BZ919249	BZ919249 CH240 650
C	114	20.2	80.8	583	9	BH042383	BH042383 RPCT-24-3
C	115	20.2	80.8	585	2	BP187346	BP187346 EST43633
C	116	20.2	80.8	599	10	CS586099	CS586099 t1gr-g88-
C	117	20.2	80.8	604	9	AQ418472	AQ418472 RPCT-11-1
C	118	20.2	80.8	615	9	CE154419	CE154419 t1gr-g88-
C	119	20.2	80.8	619	5	BO794327	BO794327 EST 3265
C	120	20.2	80.8	623	8	CV699982	CV699982 PN010E3.1
C	121	20.2	80.8	628	11	DS046629	DS046629 Oryzias 1
C	122	20.2	80.8	631	9	BH173556	BH173556 CH230-119
C	123	20.2	80.8	634	10	AG575418	AG575418 Mub muscu
C	124	20.2	80.8	638	6	CA234291	CA234291 SCCCR2309
C	125	20.2	80.8	647	9	CV503996	CV503996 70505.1 M
C	126	20.2	80.8	649	9	BH345060	BH345060 CH230-BH2
C	127	20.2	80.8	651	9	BH543412	BH543412 BOHO210TF
C	128	20.2	80.8	653	5	BM580252	BM580252 BM580252
C	129	20.2	80.8	668	10	AG054702	AG054702 Pan tlog1
C	130	20.2	80.8	676	8	CV969926	CV969926 PN008C12
C	131	20.2	80.8	684	2	BA735598	BA735598 BA473598
C	132	20.2	80.8	685	5	BX500632	BX500632 DKFPD779M
C	133	20.2	80.8	695	1	AW919295	AW919295 EST350599
C	134	20.2	80.8	697	2	BB376007	BB376007 BB376007
C	135	20.2	80.8	700	2	BE037726	BE037726 AA03C12 A
C	136	20.2	80.8	706	6	DN922421	DN922421 42641.2 C
C	137	20.2	80.8	711	10	BX199297	BX199297 Danilo rer
C	138	20.2	80.8	718	11	AO739679	AO739679 HS 5384 B
C	139	20.2	80.8	718	11	CR822540	CR822540 GR0AA52B
C	140	20.2	80.8	719	10	CE354171	CE354171 t1gr-g88-
C	141	20.2	80.8	721	9	BH722889	BH722889 BOMDT56TR
C	142	20.2	80.8	723	3	BG432218	BG432218 602496508
C	143	20.2	80.8	723	7	CV469883	CV469883 42641.1 C
C	144	20.2	80.8	724	6	CB065645	CB065645 PVRB02D08
C	145	20.2	80.8	725	7	CR851790	CR851790 12049 Stc
C	146	20.2	80.8	729	7	CR214317	CR214317 27999 S08
C	147	20.2	80.8	743	7	CR740535	CR740535 CR740535
C	148	20.2	80.8	764	9	DR719016	DR719016 AGENCOURT
C	149	20.2	80.8	779	9	BH248808	BH248808 BOGAH70TR
C	150	20.2	80.8	787	6	CB593491	CB593491 AGENCOURT

ALIGNMENTS

RESULT 1
 A0023140/c 424 bp DNA linear GSS 16-JUN-1998
 LOCUS HS_2177_A2_H07_MF_C1T Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=2177 Col=14 Row=O, genomic survey
 sequence.
 ACCESSION A0023140
 VERSION A0023140
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE 1 (bases 1 to 424)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (11), 9739-9744 (1999)
 COMMENT 1049764
 CONTACT: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2177 row: O column: 14
 Class: BAC ends
 High quality sequence stop: 424.
 Location/Qualifiers
 1..424
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="plate=2177 Col=14 Row=O"
 /sex="male"
 /clone_id="C1T Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in
 E-Coli DH10B"

ORIGIN
 Query Match 87.2%; Score 21.8; DB 9; Length 424;
 Best Local Similarity 92.0%; Pred. No. 1,76+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 AAAAAAAAAATCGACCAATCT 25
 101 AAAAAAAAAATCGCAAAATCT 77

RESULT 2
 B0929968/c 459 bp mRNA linear EST 03-JUN-2005
 LOCUS B0929968 Daphnia magna cDNA library Daphnia magna cDNA clone
 DEFINITION IG0001_0051_A09.r 3', mRNA sequence.
 ACCESSION B0929968
 VERSION B0929968
 KEYWORDS EST.
 SOURCE Daphnia magna
 ORGANISM Daphnia magna
 Arthropoda; Metazoa; Chordata; Crustacea; Branchiopoda;
 Diplostetraca; Cladocera; Anomopoda; Daphniidae; Daphnia.
 1 (bases 1 to 459)
 Watanabe,H., Tatarazako,N., Oda,S., Nishide,H., Uchiyama,I.,
 Morita,M. and Iguchi,T.
 Analysis of Expressed Sequence Tags of the water flea, Daphnia
 magna
 JOURNAL Genome (2005) In press
 COMMENT Contact: Hajime Watanabe
 Okazaki Institute for Bioscience
 National Institute of Natural Sciences
 Higashi-iyama 5-1, Myodaiji, Okazaki, Aichi, 444-8787, Japan
 Tel: 81-564-59-5237
 Fax: 81-564-59-5236
 Email: watanabe@nibb.ac.jp.
 Location/Qualifiers
 1..459
 /organism="Daphnia magna"
 /mol_type="mRNA"
 /db_xref="taxon:35525"
 /clone="IG0001_0051_A09.r"
 /sex="female"

```

ORIGIN
    /cissue_type="whole body"
    /dev_stage="adult"
    /clone_lib="Daphnia magna cDNA library"

Query Match      87.2%; Score 21.8; DB 3; Length 459;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAAATCT 25
    |||||
    277 AAAAAAAAAATCGACAAATCT 253

RESULT 3
B35642/c      492 bp      DNA      linear      GSS 17-OCT-1997
LOCUS      HS-1029-A2-G01-MR.abi CIT Human Genomic Sperm Library C Homo
DEFINITION      B35642
                B35642
                sequence.
ACCESSION      B35642.1 GI:2535011
VERSION      B35642.1 GI:2535011
KEYWORDS      GSS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Homiidae; Homo.
REFERENCE      1 (bases 1 to 492)
AUTHORS      Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
                Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.,
                Construction of a Characterized Clone Resource for Genomic
                Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
                Tagged Connectors
                Unpublished (1997)
JOURNAL      Contact: Mahairas GG, Zackrone KD, Hood L
COMMENT      University of Washington
                Seattle, WA 98195, USA
                Tel: (206) 616-8744
                Fax: (206) 685-7301
                Email: kzackrone@u.washington.edu
                Sequence Tagged Connector
                Plate: CT808 Row: M Column: 2
                Class: BAC ends
                High quality sequence scop: 492.
                Location/Qualifiers
                    1..492
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /clone="Plate=CT808 Col=2 Row=M"
                    /sex="M"
                    /clone_lib="CIT Human Genomic Sperm Library C"
                    /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
                    B-Coli DH10B"

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 492;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAAATCT 25
    |||||
    233 AAAAAAAAAATCGACAAATCT 209

RESULT 4
A2068572      495 bp      DNA      linear      GSS 30-MAR-2000
LOCUS      RPCI-23-422A23.TV RPCI-23 Mus musculus genomic clone
DEFINITION      RPCI-23-422A23, genomic survey sequence.
ACCESSION      A2068572
VERSION      A2068572.1 GI:7359824

```

```

KEYWORDS      GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 495)
AUTHORS      Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatman,S.,
                Akintet,B., Levins,M., McGann,S., Tesgaye,G., Geer,K., Krol,M., de
                Jong,P. and Fraser,C.M.
                Mouse BAC End Sequences from Library RPCI-23
                Unpublished (1999)
JOURNAL      Other_GSSes: RPCI-23-422A23.TV
COMMENT      Contact: Shaying Zhao
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: szhao@tigr.org
                Clones are derived from the mouse BAC library RPCI-23. For BAC
                library availability, please contact Pieter de Jong
                (pieter@dejong.med.buffalo.edu). Clones may be purchased from
                BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
                or from Resea ch Genetics (info@resgen.com). BAC end page:
                http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
                Plate: 422 Row: A Column: 23
                Seq primer: T7
                Class: BAC ends.
                Location/Qualifiers
                    1..495
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="RPCI-23-422A23"
                    /sex="Female"
                    /lab_host="DH10B"
                    /clone_lib="RPCI-23"
                    /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
                    EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
                    brain genomic DNA was isolated and partially digested
                    with a combination of EcoRI and EcoRI Methylase. Size
                    selected DNA was cloned into the pBACe3.6 vector at the
                    EcoRI sites. The ligation products were transformed into
                    DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 495;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACAAATCT 25
    |||||
    421 AAAAAAAAAATCGACAAATCT 445

RESULT 5
A0621055/c      507 bp      DNA      linear      GSS 16-JUN-1999
LOCUS      HS_2221_B1_C04_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION      HS_2221_B1_C04_MR CIT Approved Human Genomic Sperm Library D Homo
                sapiens genomic clone Plate=2221 Col=7 Row=F, genomic survey
                sequence.
ACCESSION      A0621055
VERSION      A0621055.1 GI:5083447
KEYWORDS      GSS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Homiidae; Homo.
REFERENCE      1 (bases 1 to 507)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

```

TITLE Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
JOURNAL Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
PUBMED Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
COMMENT 10449764
 Contact: Mahatiras GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 BAC end Web Server: http://www.htseq.washington.edu
 Plate: 2221 row F column: 7
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 507.

FEATURES
source
 Location/Qualifiers
 1..507
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=2221 Col=7 Row=F"
 /sex="male"
 /clone_id="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pbeloBAC11; BAC Clones in E-Coli DH108"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 507;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAATCGAGCAATCT 25
 157 AAAAAAAAAATCGAGCAATCT 133

RESULT 6 A2476425 574 bp DNA linear GSS 04-OCT-2000
LOCUS 1M0295H18F Mouse 10kb plasmid UGCG1M library Mus musculus genomic
DEFINITION clone UGCG1M0295H18 F, genomic survey sequence.
ACCESSION A2476425
VERSION A2476425.1 GI:10634550
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 574)
AUTHORS Jann, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weisse, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL unpublished (2000)
COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 plate: 0295 row: H column: 18
 Seq primer: CATTGTAAACGACGCGCACT
 Class: plasmid ends

FEATURES
source
 High quality sequence stop: 574.
 Location/Qualifiers
 1..574
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UGCG1M0295H18"
 /sex="male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_id="Mouse 10kb plasmid UGCG1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g1473214|g1473214|g1473214|g1473214), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 574;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAATCGAGCAATCT 25
 6 AAAAAAAAAATCGAGCAATCT 30

RESULT 7 BH512725/c 668 bp DNA linear GSS 13-DEC-2001
LOCUS BH512725
DEFINITION BOGEU92TR BOGE Brassica oleracea genomic clone BOGEU92, genomic survey sequence.
ACCESSION BH512725
VERSION BH512725.1 GI:17720815
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 668)
AUTHORS Ayle, M., Haas, B.J., Kumar, N., Wu, H., Xiao, Y., Van Aken, S., Unterbach, T.R., Wortman, J.R., White, O.R. and Town, C.D.
TITLE Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis
JOURNAL Genome Res. 15 (4), 487-495 (2005)
COMMENT Other GSSs: BOGEU92TF
 Contact: Chris Town
 TIGR
 7712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
FEATURES
source
 Location/Qualifiers
 1..668

```

/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGEU92"
/clone_1ib="BOGE"
/ncore="Vector: PHOS1, Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 668;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
    |||||
Db 536 AAAAAAAAAATCTCTGACAAATCT 512

RESULT 8
AG099199          682 bp  DNA      linear  GSS 03-NOV-2001
LOCUS             Pan troglodytes DNA, clone: PTB-101G12.F, genomic survey sequence.
ACCESSION         AG099199
VERSION           AG099199.1 GI:16719716
KEYWORDS          GSS.
SOURCE            Pan troglodytes (chimpanzee)
ORGANISM          Pan troglodytes
AUTHORS           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                  Homnidae; Pan.
REFERENCE
  1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
    Toroki,Y., Watanabe,H. and Sakaki,Y.
    BAC end sequences of library PTB
    Unpublished
  2 (bases 1 to 682)
    Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
    Toroki,Y., Watanabe,H. and Sakaki,Y.
    Direct Submission
    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
    and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
    1-7-22 Suenitro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
    (E-mail:chimpe@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
    Tel:81-45-503-9111, Fax:81-45-503-9170)
    Clones are derived from the chimpanzee BAC library PTB This BAC end
    was generated during the Rad process and may have higher chance of
    clone tracking errors.
    PRIMERS
    Sequencing: -21M13
    LIBRARY
    Vector : PKS145
    R.Site 1 : SacI
    R.Site 2 : SacI.
    Location/Qualifiers
      1. .682
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /clone="PTB-101G12.F"
        /sex="male"
        /cell_type="lymphoblast"
        /clone_1ib="PTB Chimpanzee Male BAC library"

ORIGIN
Query Match      87.2%; Score 21.8; DB 10; Length 682;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
    |||||
Db 274 AAAAAAAAAATCGCAAAATCT 298

```

```

RESULT 9
BH248096/c       795 bp  DNA      linear  GSS 26-NOV-2001
LOCUS            BOCAT53TF BOGA Brassica oleracea genomic clone BOCAT53, genomic
DEFINITION       survey sequence.
ACCESSION        BH248096
VERSION          BH248096.1 GI:17068872
KEYWORDS         GSS.
SOURCE           Brassica oleracea
ORGANISM         Brassica oleracea
AUTHORS          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
                  rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
  1 (bases 1 to 795)
    Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
    Uteerback,T.R., Mortman,J.R., White,O.R. and Town,C.D.
    Whole genome shotgun sequencing of Brassica oleracea and its
    application to gene discovery and annotation in Arabidopsis
    Genome Res. 15 (4), 487-495 (2005)
    Other GSSs: BOCAT53TR
    Contact: Chris Town
    TIGR Medical Center Drive, Rockville, MD 20850, USA.
    9712
    Tel: 301-838-3523
    Fax: 301-838-0208
    Email: cdtown@tigr.org
    DNA is from a doubled haploid provided by Tom Osborn.
    Seq primer: TP
    Class: sheared ends.
    Location/Qualifiers
      1. .795
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="TO1000DH3"
        /db_xref="taxon:3712"
        /clone="BOCAT53"
        /clone_1ib="BOGA"
        /ncore="Vector: PHOS1, Site_1: BstXI; 2-3 kb sheared
        genomic DNA inserted into PHOS1 using BstXI linkers"

ORIGIN
Query Match      87.2%; Score 21.8; DB 9; Length 795;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
    |||||
Db 255 AAAAAAAAAATCGCAAAATCT 231

RESULT 10
BU509223/c       898 bp  mRNA      linear  EST 12-SEP-2002
LOCUS            AGENCOURT 10095473 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6503781
DEFINITION       5', mRNA sequence.
ACCESSION        BU509223
VERSION          BU509223.1 GI:22815456
KEYWORDS         EST.
SOURCE           Homo sapiens (human)
ORGANISM         Homo sapiens
AUTHORS          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                  Homnidae; Homo
REFERENCE
  1 (bases 1 to 898)
    NIH-MGC http://mgi.nci.nih.gov/.
    National Institutes of Health, Mammalian Gene Collection (MGC)
    Unpublished (1999)
    Contact: Robert Strausberg, Ph.D.
    Email: cga@bbs-remail.nih.gov
    Tissue Procurement: ATCC
    cDNA Library Preparation: Life Technologies, Inc.

```

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
 Plate: LLM14061 row: P column: 22
 High quality sequence stop: 585.

Location/Qualifiers
 1..898

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6503781"
 /tissue_type="leiomyosarcoma"
 /lab_host="RDH10B (phage-resistant)"
 /clone_lib="NIH_MGC_71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."

ORIGIN

Query Match 87.2%; Score 21.8; DB 5; Length 898;
 Best Local Similarity 92.0%; Pred. No. 1.7e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACAGCAATCT 25
 Db 759 AAAAAAAAAATCGCAAAAAATCT 735

RESULT 11
 CE581913 511 bp DNA linear GSS 28-SEP-2003
 LOCUS tigr-gss-dog-1700036273250 Dog Library Canis familiaris genomic,
 DEFINITION genomic survey sequence.
 ACCESSION CE581913
 VERSION CE581913.1 GI:36898694
 KEYWORDS GSS.
 SOURCE
 ORGANISM Canis familiaris (dog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 Canis.
 1 (bases 1 to 511)
 Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
 Venter,J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 14512627

COMMENT

Contact: Kirkness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirkness@tigr.org
 Class: shotgun.

FEATURES

source

1..511
 Location/Qualifiers
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BclXI; Libraries were prepared from
 peripheral blood"

ORIGIN

Query Match 85.6%; Score 21.4; DB 10; Length 511;
 Best Local Similarity 95.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACAGCAAT 23
 Db 416 AAAAAAAAAATGCGACAGCAAT 394

RESULT 12
 CE372285 614 bp DNA linear GSS 27-SEP-2003
 LOCUS tigr-gss-dog-17000362172049 Dog Library Canis familiaris genomic,
 DEFINITION genomic survey sequence.
 ACCESSION CE372285
 VERSION CE372285.1 GI:36595601
 KEYWORDS GSS.
 SOURCE
 ORGANISM Canis familiaris (dog)
 Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 Canis.
 1 (bases 1 to 614)

REFERENCE
 AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
 Venter,J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 14512627

COMMENT
 Contact: Kirkness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirkness@tigr.org
 Class: shotgun.

FEATURES
 source
 Location/Qualifiers
 1..614
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BclXI; Libraries were prepared from
 peripheral blood"

ORIGIN

Query Match 85.6%; Score 21.4; DB 10; Length 614;
 Best Local Similarity 95.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACAGCAAT 23
 Db 403 AAAAAAAAAATGCGACAGCAAT 425

RESULT 13
 CE781479 706 bp DNA linear GSS 30-SEP-2003
 LOCUS tigr-gss-dog-17000330656093 Dog Library Canis familiaris genomic,
 DEFINITION genomic survey sequence.
 ACCESSION CE781479
 VERSION CE781479.1 GI:37122242
 KEYWORDS GSS.
 SOURCE
 ORGANISM Canis familiaris (dog)
 Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 Canis.
 1 (bases 1 to 706)

REFERENCE
 AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
 Venter,J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 14512627

COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Classes: shotgun,
Location/Qualifiers

FEATURES
source
1..706
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poolle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN

Query Match 85.6%; Score 21.4; DB 10; Length 706;
Best Local Similarity 95.7%; Pred. No. 2.3e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATC 23
Db 446 AAAAAAAAAATGACGACAAAT 468

RESULT 14
BJ683132 199 bp mRNA linear EST 25-JAN-2005
LOCUS BJ683132 HCEST library Haplochromis chilotas cDNA clone no77c01,
DEFINITION mRNA sequence.
BJ683132
ACCESSION BJ683132.1 GI:46526253
VERSION
KEYWORDS
SOURCE Haplochromis chilotas
ORGANISM Haplochromis chilotas
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;
Labroidae; Cichlidae; African cichlids; Pseudocrenilabrinae;
Haplochromini; Haplochromis.
1 (bases 1 to 199)
Watanabe,M., Kobayashi,N., Shin-I,T., Horiike,T., Tateno,Y.,
Kohara,Y. and Okada,N.
Extensive analysis of ORF sequences from two different cichlid
species in Lake Victoria provides molecular evidence for a recent
radiation event of the Victoria species flock: identity of EST
sequences between Haplochromis chilotas and Haplochromis sp.
JOURNAL Gene 343 (2), 263-269 (2004)
PUBMED 15588581
COMMENT Contact: Tadaa Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..199
/organism="Haplochromis chilotas"
/mol_type="mRNA"
/db_xref="taxon:257977"
/clone="no77c01"
/tissue="whole embryo"
/dev_stage="embryonic 7hr"
/lab_host="E.coli, XL1 blue"
/clone_lib="Sea urchin embryo 7hr cleavage stage cDNA
library MPMGP538"
/note="vector: pSPori1; Site 1: NotI; Site 2: SalI; Random
primed and directionally cloned in pSPori1 vector using a
NotI (5'-pGACTGATGCTGATGCGAGCGCGCC (T)15-3' and a
SalI 5'-TCGACCCAGCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 83.2%; Score 20.8; DB 3; Length 199;
Best Local Similarity 91.7%; Pred. No. 4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATC 24
Db 176 AAAAAAAAAATCGCTGAAAAATC 199

RESULT 15
CD290160 382 bp mRNA linear EST 16-SEP-2003
LOCUS CD290160
DEFINITION Strp538.003119 Sea urchin embryo 7hr cleavage stage cDNA library
MPMGP538 Strongylocentrotus purpuratus cDNA clone
CALTP538J072.MPI_538_207 3', mRNA sequence.
CD290160
ACCESSION CD290160.1 GI:34741237
VERSION
KEYWORDS
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinoidea; Echinodermata; Echinodermata;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 382)
Poustka,A.J., Groth,D., Hennig,S., Thamm,S., Cameron,A., Beck,A.,
Reinhardt,R., Herwig,R., Panopoulou,G. and Lehrach,H.
Generation, annotation, evolutionary analysis, and database
integration of 20,000 unique sea urchin EST clusters
Genome Res. 13 (12), 2736-2746 (2003)
Contact: Poustka AJ
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: poustka@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONF) to reduce sequencing redundancy. According to the ONF
procedure, clones that display the same hybridisation matrix with a
battery of 200 6mer oligonucleotides are grouped into clusters. One
clone per ONF cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well as the coordinates of
the other clones assigned to the same ONF cluster as the clone from
which the above EST is generated is available at the sea urchin
project web site at: http://www.molgen.mpg.de/ag_seaurchin/. cDNA
clones and filters are distributed via the Resource Center/Primary
Database of the German Human Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTATGCTTCGGCTCG 3' (M138SP) 5'-seq
BACKWARD: 5' GCTATTACGCCAGCTGCGAAGGCGATG 3' (M13FSF) 3'-seq
Seq primer: 5' GCTATTACGCCAGCTGCGAAGGCGGATG 3' (M13FSF)
High quality sequence stop: 382.
Location/Qualifiers

FEATURES
source
1..382
/organism="Strongylocentrotus purpuratus"
/mol_type="mRNA"
/db_xref="taxon:7668"
/clone="CALTP538J072.MPI_538_207"
/tissue="whole embryo"
/dev_stage="embryonic 7hr"
/lab_host="E.coli, XL1 blue"
/clone_lib="Sea urchin embryo 7hr cleavage stage cDNA
library MPMGP538"
/note="vector: pSPori1; Site 1: NotI; Site 2: SalI; Random
primed and directionally cloned in pSPori1 vector using a
NotI (5'-pGACTGATGCTGATGCGAGCGCGCC (T)15-3' and a
SalI 5'-TCGACCCAGCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 382;
Best Local Similarity 91.7%; Pred. No. 3.9e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATC 24
|||||

Db 73 AAAAAAAAAATCGATTCAATC 96

RESULT 16
LOCUS B0632287/c 477 bp mRNA linear EST 02-JUL-2002
DEFINITION 1125a10.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6031002 3', mRNA sequence.

ACCESSION B0632287
KEYWORDS B0632287.1 GI:21683805
VERSION EST.
KEYWORDS B0632287.1 GI:21683805
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 477)
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradowol, G., Clifton, S., Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Bliseth, A., Schmitt, A., Theising, B., Ritter, E., Konko, T., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarishevili, R., Williams, T., Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 381.
Location/Qualifiers
1. 477
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6031002"
/cissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_id="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permut Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@ingate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 477;
Best Local Similarity 88.0%; Pred. No. 3.8e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 AAAAAAAAAATCGACGAATCT 25
49 AAAAAAAAAATCCTGACAAAGCT 25

RESULT 17
LOCUS CL761036/c 599 bp DNA linear GSS 27-JUL-2004
DEFINITION OR_BBa0130L11.r OR_BBa Oryza nivara genomic clone OR_BBa0130L11 3', genomic survey sequence.

ACCESSION CL761036
VERSION CL761036.1 GI:50717721
KEYWORDS GSS.
SOURCE Oryza nivara
ORGANISM Oryza nivara
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriactoidae; Oryzaceae; Oryza.
1 (bases 1 to 599)
Kim, H., Yu, Y., Stum, D., Yost, D., Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Hatfield, J., Soderlund, C. and Wing, R.
OMAP Project
Unpublished (2004)
Contact: Rod A. Wing
Arizona Genomes Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: twing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Insert length: 161
Plate: 0130 row: 1 Column: 11
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
Location/Qualifiers
1. 599
/organism="Oryza nivara"
/mol_type="genomic DNA"
/db_xref="taxon:4536"
/clone="OR_BBa0130L11"
/cissue_type="young leaves"
/lab_host="DH10B-T1 phage resistant"
/clone_id="OR_BBa"
/note="Vector: pACIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 599;
Best Local Similarity 91.7%; Pred. No. 3.8e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2 AAAAAAAAAATCGACGAATCT 25
184 AAAAAAAAAATCTCGAAGAAATCT 161

RESULT 18
LOCUS CN006484 605 bp mRNA linear EST 26-MAR-2004
DEFINITION CSECS134E04.CELu0001 CabSau Cell Culture (CELu0001) Vitis vinifera
CNuA clone CSECS134E04 3', mRNA sequence.
CN006484
CN006484.1 GI:45770632
EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; Vitaceae; Vitis.
1 (bases 1 to 605)
Iocco, P., Davies, C. and Thomas, M.R.
Expressed sequence tags from the grapevine cultivar Cabernet Sauvignon (2004)
Unpublished (2004)
Contact: Mark R. Thomas
CSIRO Plant Industry
CSIRO
PO Box 350, Glen Osmond, SA, 5064, Australia
Tel: 61 8 83038600
Fax: 61 8 83038601
Email: Mark.R.Thomas@csiro.au

Seq primer: TAATGCACTCACTATAGG (T7)
POLYA=yes.

FEATURES

Source

Location/Qualifiers
1..605

/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSBCS134E04"
/sex="Hermaphrodite"
/cell_type="Cell Suspension Culture"
/clone_lib="Cabsau Cell Culture (CELn0001)"
/note="Vector: pTRIPLEX2. A cDNA library from a Cabernet Sauvignon cell suspension culture."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 605;
Best Local Similarity 91.7%; Pred. No. 3.8e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACGACCAATC 24

Db 535 AAAAAAAAAATCGACGACCAATC 558

RESULT 19

CV709331

633 bp mRNA linear EST 03-NOV-2004

LOCUS UCRPT01.001210.7 r Poncirus trifoliata CTV-challenged cDNA library -

DEFINITION AG12 Poncirus trifoliata cDNA clone PT_6BA001210.7, mRNA sequence.

ACCESSION CV709331

VERSION CV709331.1 GI:55291699

KEYWORDS EST.

SOURCE Poncirus trifoliata

ORGANISM Poncirus trifoliata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

1 (bases 1 to 633)

Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,

Manamaker, S., Kim, H.R., Kudrna, D. and Stum, D., Wisconsin, M.,

Wing, R.

Development of EST Resources and New Genetic Markers for California

Citrus - Poncirus trifoliata CTV-challenged phloem - AG12

Unpublished (2004)

Contact: Mikeal Roose

Department of Botany & Plant Sciences, University of California

Riverside, CA, 92521-0124, USA

Tel: 9097874137

Fax: 9097874437

Email: mikeal.roose@ucr.edu

Seq primer: T3.

Location/Qualifiers

1..633

/organism="Poncirus trifoliata"

/mol_type="mRNA"

/cultivar="Pomeroi OP"

/db_xref="taxon:37690"

/clone="PT_6BA001210.7"

/tissue_type="Phloem"

/dev_stage="10 - 30 cm shoots"

/lab_host="E. coli TUC121"

/clone_lib="Poncirus trifoliata CTV-challenged cDNA

library - AG12"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site 1: BcORI, Site 2: XhoI; Plants were grown in the

greenhouse at University of California, Riverside. The

seedling of Poncirus trifoliata cv Pomeroi that was

selected as homozygous for the CTV resistance gene. The

rootstock was sweet orange infected with citrus tristeza

virus (CTV) isolate 7514 over 1 year before sampling (CTV

infects sweet orange, but not genotypes carrying the CTV

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 633;
Best Local Similarity 91.7%; Pred. No. 3.8e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACGACCAATC 24

Db 57 AAAAAAAAAATCGACGACCAATC 80

RESULT 20

CV643942

675 bp mRNA linear EST 18-JAN-2005

LOCUS UCRPT02.70G01.9 Poncirus trifoliata Roots with Iron Deficiency -

DEFINITION UCRPT02-UCR1 Poncirus trifoliata cDNA clone UCRPT02-70G01-N1-5-9,

mRNA sequence.

ACCESSION CV643942

VERSION CV643942.1 GI:57878771

KEYWORDS EST.

SOURCE Poncirus trifoliata

ORGANISM Poncirus trifoliata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

1 (bases 1 to 675)

Fenton, R.D., Manamaker, S., Landry, B., Hubert, N., Laforest, M.,

Landry, J. and Ligonde, A.

Development of EST Resources and New Genetic Markers for California

Citrus - Poncirus trifoliata Roots with Iron Deficiency -

UCRPT02-UCR1

Unpublished (2005)

Contact: Timothy J. Close

Department of Botany & Plant Sciences

University of California

Riverside, CA 92521-0124, USA

Tel: 909-787-3318

Fax: 909-787-4437

Email: timothy.close@ucr.edu

Seq primer: T3.

Location/Qualifiers

1..675

/organism="Poncirus trifoliata"

/mol_type="mRNA"

/cultivar="Pomeroi"

/db_xref="taxon:37690"

/clone="UCRPT02-70G01-N1-5-9"

/tissue_type="root"

/dev_stage="seedling"

/lab_host="E. coli TUC121"

/clone_lib="Poncirus trifoliata Roots with Iron Deficiency

- UCRPT02-UCR1"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;

resistance gene). Shoots 10-30 cm long were harvested in October 2000, and the green phloem (dark) was removed and frozen quickly in dry ice. Total RNA was extracted using TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of phagescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Kim, Kudrna, Stum, Wisconsin, Wing). Chromatogram files were downloaded to UC Riverside (Close), then processed at UC Riverside (Manamaker) using the HarvEST pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to Genbank."

Site 1: Ecor1; Site 2: Xho1; plant materials were prepared by Federici (Roose Lab) with advice from Parker. Seedlings 138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO₂ buffering. This solution was replaced on July 6, 2004. Roots from three plants were sampled on June 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandel and Fenton (Close lab) purified RNA using TRIzol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FFP by Close, then processed by Mananaker (Close lab) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Mananaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 675;
Best Local Similarity 91.7%; Pred. No. 3.8e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACACAATC 24
Db 48 AAAAAAAAAATCGACACAATC 71

RESULT 21

CN980172 720 bp mRNA linear EST 08-JUN-2004
LOCUS 31503 125 104 H09 Fundulus heteroclitus Liver Fundulus heteroclitus
DEFINITION cDNA similar to Cysteratin precursor, mRNA sequence.
ACCESSION CN980172
VERSION CN980172.1 GI:48461763
KEYWORDS EST.
SOURCE Fundulus heteroclitus (killifish)

ORGANISM

Fundulus heteroclitus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

4600 Rickenbacker Causeway, Miami, FL 33149-1098 USA
Tel: 305 361 4121
Email: dcrawford@rmas.miami.edu
Database Web Interface
http://genomics.rmas.miami.edu/funnybase/super_craw3/
Plate: 104 row: H column: 9.
Location/Qualifiers
1..720
/organism="Fundulus heteroclitus"
/mol_type="mRNA"
/db_xref="taxon:8078"
/feature_type="Liver"
/clone_id="Fundulus Heteroclitus Liver"
/note="Organ: Liver"

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 720;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACACAATC 24
Db 675 AAAAAAAAAATCGACACAATC 698

RESULT 22

CX670454 729 bp mRNA linear EST 19-JAN-2005
LOCUS UCRCP01.050 H12.T7 Swingle citrinello nematode-challenged root cDNA
DEFINITION library_ UCRCP01.050.T7_H12 mRNA sequence.
ACCESSION CX670454
VERSION CX670454.1 GI:57929549
KEYWORDS EST.
SOURCE Citrus x paradisi x Poncirus trifoliata
ORGANISM Citrus x paradisi x Poncirus trifoliata
REFERENCE Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Sapindales; Rutaceae; Citrus x Poncirus.
1 (bases 1 to 729)
Fenton, R.D., Mananaker, S., Choi, Y.G. and Kingan, T.
Development of EST Resources and New Genetic Markers for California Citrus - Swingle citrinello nematode-challenged roots - UCRCP01 Unpublished (2005)
Contact: Timothy J. Close
Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timothy.close@ucr.edu
Seq primer: T7.
Location/Qualifiers
1..729
/organism="Citrus x paradisi x Poncirus trifoliata"
/mol_type="mRNA"
/cultiivar="Swingle"
/db_xref="taxon:309804"

FEATURES

source

/clone="UCRCP01_050_T7_H12"
 /tissue_type="root"
 /dev_stage="seedling"
 /lab_host="E. coli TUC121"
 /clone_lib="Swingle citrumeelo nematode-challenged root
 cDNA library - UCRCP01"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; twelve seedlings were
 transplanted from cones to 4 inch pots of sand mix, then
 grown in greenhouses (Darsow, Becker lab). Citrus
 nematodes (Tylenchulus semipenetrans) were extracted from
 contaminated soil from the UC Riverside Citrus Research
 Center then suspended in water and added to pots (Darsow).
 Growing roots were collected by clipping them away from
 the old root ball 24 hr and 4 days after inoculation
 (Federict). Total RNA was extracted using Concert Plant
 RNA Reagent (Invitrogen) followed by RNeasy (Qiagen), an
 equal amount of RNA was pooled from each of the two
 collection times, poly(A) RNA was purified using Oligotex
 (Qiagen), a cDNA library was made using a cDNA synthesis
 kit (Stratagene), then 0.45 million primary lambda cDNA
 clones were in vivo excised to give a population of
 Bluescript SK(-) phagemids (Fenton, Close lab). Phagemids
 were plated, plasmid DNA purified, cDNA clones archived,
 and DNA sequences determined bi-directionally using an
 ABI3730 at the University of California Riverside
 Institute of Integrative Genome Biology Genomics Core
 Instrumentation Facility (Choi, Kingan). Chromatogram
 files were downloaded by FTP by Close, then processed by
 Manamaker (Close lab) using the HarVEST pipeline
 (http://harvest.ucr.edu) to remove vector and cloning
 oligo sequences and various contaminants, and to trim to a
 high quality region. Sequences that retained a phred 17
 region of at least 100 bases were assembled, then chimeras
 were removed following manual inspection of assemblies
 (Close, Roose, Manamaker). Sequences that survived all
 removal steps were submitted to GenBank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 729;
 Best Local Similarity 91.7%; Pred. No. 3.7e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACACAATC 24
 |||||
 Db 700 AAAAAAAAAATCACAGAAATC 677

RESULT 23
 CX543595/c 731 bp mRNA linear EST 12-JAN-2005
 LOCUS UCRPT01_5_007_H06_T7 Poncirus trifoliata CTV-challenged cDNA
 DEFINITION library - UCRPT01-UCR2 Poncirus trifoliata cDNA clone
 UCRPT01_007_T7_H06, mRNA sequence.
 ACCESSION CX543595
 VERSION CX543595.1 GI:57570620
 SOURCE EST.
 ORGANISM Poncirus trifoliata
 Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eustosids II; Sapindales; Rutaceae; Poncirus.
 1 (bases 1 to 731)
 Roose, M.L., Ye, X., Federict, C.F., Close, T.J., Fenton, R.D.,
 Manamaker, S., Choi, Y. and Kingan, J.
 Development of EST Resources and New Genetic Markers for California
 Citrus - Poncirus trifoliata CTV-challenged phloem - UCRPT01-UCR2
 Unpublished (2004)
 CONTACT: Mikel Rose
 Department of Botany & Plant Sciences, University of California
 Riverside, CA, 92521-0124, USA
 Tel: 9097874137
 Fax: 9097874437

Email: mikel.rose@ucr.edu
 Seq primer: T7.
 FEATURES
 source
 location/Qualifiers
 1..731
 /organism="Poncirus trifoliata"
 /mol_type="mRNA"
 /cultivar="Pomeroy OP"
 /db_xref="taxon:37690"
 /clone="UCRPT01_007_T7_H06"
 /tissue_type="bark (with phloem)"
 /dev_stage="10 - 30 cm shoots"
 /lab_host="E. coli TUC121"
 /clone_lib="Poncirus trifoliata CTV-challenged cDNA
 library - UCRPT01-UCR2"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
 greenhouse at University of California, Riverside. The
 action was an open-pollinated (very probably selfed)
 seedling of Poncirus trifoliata cv Pomeroy that was
 selected as homozygous for the CTV resistance gene. The
 rootstock was sweet orange infected with citrus tristeza
 virus (CTV) isolate TS14 over 1 year before sampling (CTV
 infects sweet orange, but not genotypes carrying the CTV
 resistance gene. Shoots 10-30 cm long were harvested in
 October 2000, and the green phloem (bark) was removed and
 frozen quickly in dry ice. Total RNA was extracted using
 Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA
 library was made, and 0.5 million primary lambda cDNA
 clones were in vivo excised to give a population of
 Bluescript SK(-) phagemids. All steps to this point were
 performed in the M.L. Roose lab at the University of
 California, Riverside by X. Ye. Phagemids were plated,
 plasmid DNA purified, cDNA clones archived, and DNA
 sequences determined bi-directionally using an ABI3730 at
 the University of California Riverside Institute of
 Integrative Genome Biology Genomics Core Instrumentation
 Facility, (Choi, Kingan). Chromatogram files were
 downloaded by FTP by Close, then processed by Manamaker
 (Close lab) using the HarVEST pipeline
 (http://harvest.ucr.edu) to remove vector and cloning
 oligo sequences and various contaminants, and to trim to a
 high quality region. Sequences that retained a phred 17
 region of at least 100 bases were assembled, then chimeras
 were removed following manual inspection of assemblies
 (Close, Roose, Manamaker). Sequences that survived all
 removal steps were submitted to GenBank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 731;
 Best Local Similarity 91.7%; Pred. No. 3.7e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACACAATC 24
 |||||
 Db 731 AAAAAAAAAATCACAGAAATC 708

RESULT 24
 CX666775 732 bp mRNA linear EST 19-JAN-2005
 LOCUS UCRCP01_025_C12_T7 Swingle citrumeelo nematode-challenged root cDNA
 DEFINITION library - UCRCP01 Citrus x paradisi x Poncirus trifoliata cDNA
 clone UCRCP01_025_T7_C12, mRNA sequence.
 ACCESSION CX666775
 VERSION CX666775.1 GI:57925322
 SOURCE EST.
 ORGANISM Citrus x paradisi x Poncirus trifoliata
 Citrus x paradisi x Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eustosids II; Sapindales; Rutaceae; Citrus x Poncirus.
 1 (bases 1 to 732)
 Close, T.J., Roose, M.L., Becker, O., Darsow, J., Federict, C.F.,

RESULT 26
 CX666208/c
 LOCUS
 DEFINITION CX666208 739 bp mRNA linear EST 19-JAN-2005
 UCRCP01_021_D12_T7 Swingle citrumeelo nematode-challenged root cDNA
 library - UCRCP01 Citrus x paradisi x Poncirus trifoliata cDNA
 clone UCRCP01_021_T7_D12, mRNA sequence.
 CX666208
 CX666208.1 GI:57924642
 EST.
 ORGANISM
 SOURCE Citrus x paradisi x Poncirus trifoliata
 Citrus x paradisi x Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Sapindales; Rutaceae; Citrus x Poncirus.
 1 (bases 1 to 739)
 Close,T.J., Roose,M.L., Becker,O., Darsow,J., Federici,C.F.,
 Fenton,R.D., Manamaker,S., Choi,Y.G and Kington,T.
 Development of EST Resources and New Genetic Markers for California
 Citrus - Swingle citrumeelo nematode-challenged roots - UCRCP01
 Unpublished (2005)
 CONTACT: Timothy J. Close
 Department of Botany & Plant Sciences
 University of California
 Riverside, CA 92521-0124, USA
 Tel: 909-787-3318
 Fax: 909-787-4437
 Email: timothy.close@ucr.edu
 Seq primer: T7.
 Location/Qualifiers
 1..739
 /organism="Citrus x paradisi x Poncirus trifoliata"
 /mol_type="mRNA"
 /cultivar="Swingle"
 /db_xref="taxon:309804"
 /clone="UCRCP01_021_T7_D12"
 /tissue_type="root"
 /dev_stage="seedling"
 /lab_host="E. coli TUC121"
 /clone_lib="Swingle citrumeelo nematode-challenged root
 cDNA library - UCRCP01"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Twelve seedlings were
 transplanted from cones to 4 inch pots of sand mix, then
 grown in greenhouses (Darsow, Becker lab). Citrus
 nematodes (Tylenchulus semipenetrans) were extracted from
 contaminated soil from the UC Riverside Citrus Research
 Center then suspended in water and added to pots (Darsow).
 Growing roots were collected by clipping them away from
 the old root ball 24 hr and 4 days after inoculation
 (Federici). Total RNA was extracted using Concert Plant
 RNA Reagent (Invitrogen) followed by RNeasy (Qiagen), an
 equal amount of RNA was pooled from each of the two
 collection times, poly(A) RNA was purified using Oligotex
 kit (Stratagene), then 0.45 million primary lambda cDNA
 clones were in vivo excised to give a population of
 pluescript SK(-) phagemids (Fenton, Close lab). Phagemids
 were plated, plasmid DNA purified, cDNA clones archived,
 and DNA sequences determined bi-directionally using an
 ABI3730 at the University of California Riverside
 Institute of Integrative Genome Biology Genomics Core
 Instrumentation Facility. (Choi, Kington). Chromatogram
 files were downloaded by FTP by Close, then processed by
 Manamaker (Close lab) using the Harvest pipeline
 (http://harvest.ucr.edu) to remove vector and cloning
 oligo sequences and various contaminants, and to trim to a
 high quality region. Sequences that retained a phred 17
 region of at least 100 bases were assembled, then chimeras
 were removed following manual inspection of assemblies
 (Close, Roose, Manamaker). Sequences that survived all
 removal steps were submitted to GenBank."

ORIGIN
 Query Match 83.2%; Score 20.8; DB 8; Length 739;
 Best Local Similarity 91.7%; Pred. No. 3.7e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAATCGACGAAATC 24
 Db 701 AAAAAAAAAATCAGGAAATC 678
 RESULT 27
 CX666774/c
 LOCUS
 DEFINITION CX666774 739 bp mRNA linear EST 19-JAN-2005
 UCRCP01_025_C12_T3 Swingle citrumeelo nematode-challenged root cDNA
 library - UCRCP01 Citrus x paradisi x Poncirus trifoliata cDNA
 clone UCRCP01_025_T3_C12, mRNA sequence.
 CX666774
 CX666774.1 GI:57925319
 EST.
 ORGANISM
 SOURCE Citrus x paradisi x Poncirus trifoliata
 Citrus x paradisi x Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Sapindales; Rutaceae; Citrus x Poncirus.
 1 (bases 1 to 739)
 Close,T.J., Roose,M.L., Becker,O., Darsow,J., Federici,C.F.,
 Fenton,R.D., Manamaker,S., Choi,Y.G and Kington,T.
 Development of EST Resources and New Genetic Markers for California
 Citrus - Swingle citrumeelo nematode-challenged roots - UCRCP01
 Unpublished (2005)
 CONTACT: Timothy J. Close
 Department of Botany & Plant Sciences
 University of California
 Riverside, CA 92521-0124, USA
 Tel: 909-787-3318
 Fax: 909-787-4437
 Email: timothy.close@ucr.edu
 Seq primer: T3.
 Location/Qualifiers
 1..739
 /organism="Citrus x paradisi x Poncirus trifoliata"
 /mol_type="mRNA"
 /cultivar="Swingle"
 /db_xref="taxon:309804"
 /clone="UCRCP01_025_T3_C12"
 /tissue_type="root"
 /dev_stage="seedling"
 /lab_host="E. coli TUC121"
 /clone_lib="Swingle citrumeelo nematode-challenged root
 cDNA library - UCRCP01"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Twelve seedlings were
 transplanted from cones to 4 inch pots of sand mix, then
 grown in greenhouses (Darsow, Becker lab). Citrus
 nematodes (Tylenchulus semipenetrans) were extracted from
 contaminated soil from the UC Riverside Citrus Research
 Center then suspended in water and added to pots (Darsow).
 Growing roots were collected by clipping them away from
 the old root ball 24 hr and 4 days after inoculation
 (Federici). Total RNA was extracted using Concert Plant
 RNA Reagent (Invitrogen) followed by RNeasy (Qiagen), an
 equal amount of RNA was pooled from each of the two
 collection times, poly(A) RNA was purified using Oligotex
 kit (Stratagene), then 0.45 million primary lambda cDNA
 clones were in vivo excised to give a population of
 pluescript SK(-) phagemids (Fenton, Close lab). Phagemids
 were plated, plasmid DNA purified, cDNA clones archived,
 and DNA sequences determined bi-directionally using an
 ABI3730 at the University of California Riverside
 Institute of Integrative Genome Biology Genomics Core
 Instrumentation Facility. (Choi, Kington). Chromatogram
 files were downloaded by FTP by Close, then processed by
 Manamaker (Close lab) using the Harvest pipeline
 (http://harvest.ucr.edu) to remove vector and cloning
 oligo sequences and various contaminants, and to trim to a
 high quality region. Sequences that retained a phred 17
 region of at least 100 bases were assembled, then chimeras
 were removed following manual inspection of assemblies
 (Close, Roose, Manamaker). Sequences that survived all
 removal steps were submitted to GenBank."

FEATURES

source

files were downloaded by FTP by Close, then processed by Manamaker (Close lab) using the Harvest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Manamaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 739;
 Beet Local Similarity 91.7%; Pred. No. 3.7e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAGCAAAATC 24
 |||||
 685 AAAAAAAAAATCGCAGCAAAATC 662

Db

RESULT 28

LOCUS

CK543594 743 bp mRNA linear EST 12-JAN-2005
 DEFINITION UCRPT01_5_007_H06_T3 Poncirus trifoliata CTV-challenged CDNA

library - UCRPT01-UCR2 Poncirus trifoliata CDNA clone
 UCRPT01_007_T3_H06, mRNA sequence.

ACCESSION

CK543594
 VERSION
 KEYWORDS
 EST.

SOURCE

ORGANISM

Poncirus trifoliata
 Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosidae; eurosids II; Sapindales; Rutaceae; Poncirus.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1 (bases 1 to 743)
 Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,
 Manamaker, S., Choi, Y. and Kingan, T.
 Development of EST Resources and New Genetic Markers for California
 Citrus - Poncirus trifoliata CTV-challenged phloem - UCRPT01-UCR2
 Unpublished (2004)
 Contact: Mikel Roose
 Department of Botany & Plant Sciences, University of California
 Riverside, CA, 92521-0124, USA
 Tel: 9097874137
 Fax: 9097874437
 Email: mikel.roose@ucr.edu
 Seq primer: T3.
 Location/Qualifiers

FEATURES

source

1..743
 /organism="Poncirus trifoliata"
 /mol_type="mRNA"
 /cultivar="Pomeroy OP"
 /db_xref="taxon:37690"
 /clone="UCRPT01_007_T3_H06"
 /tissue_type="bark (with phloem)"
 /dev_stage="10 - 30 cm shoots"
 /lab_host="E. coli TUC121"
 /clone_lib="Poncirus trifoliata CTV-challenged CDNA
 library - UCRPT01-UCR2"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI; Site 2: XhoI; plants were grown in the
 greenhouse at University of California, Riverside. The
 action was an open-pollinated (very probably selfed)
 seedling of Poncirus trifoliata cv Pomeroy that was
 selected as homozygous for the CTV resistance gene. The
 rootstock was sweet orange infected with citrus tristeza
 virus (CTV) isolate T514 over 1 year before sampling (CTV
 infects sweet orange, but not genotypes carrying the CTV
 resistance gene. Shoots 10-30 cm long were harvested in
 October 2000, and the green phloem (bark) was removed and
 frozen quickly in dry ice. Total RNA was extracted using
 Trizol reagent (Gibco). Poly(A) RNA was purified, a CDNA
 library was made, and 0.5 million primary lambda cDNA

clones were in vivo excised to give a population of
 pBluescript SK(-) phagemids. All steps to this point were
 performed in the M. Roose lab at the University of
 California, Riverside by X. Ye. Phagemids were plated,
 plasmid DNA purified, cDNA clones archived, and DNA
 sequences determined bi-directionally using an ABI3730 at
 the University of California Riverside Institute of
 Integrative Genome Biology Genomics Core Instrumentation
 Facility, (Choi, Kingan). Chromatogram files were
 downloaded by FTP by Close, then processed by Manamaker
 (Close lab) using the Harvest pipeline
 (<http://harvest.ucr.edu>) to remove vector and cloning
 oligo sequences and various contaminants, and to trim to a
 high quality region. Sequences that retained a phred 17
 region of at least 100 bases were assembled, then chimeras
 were removed following manual inspection of assemblies
 (Close, Roose, Manamaker). Sequences that survived all
 removal steps were submitted to GenBank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 743;
 Beet Local Similarity 91.7%; Pred. No. 3.7e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAGCAAAATC 24
 |||||
 36 AAAAAAAAAATCGCAGCAAAATC 59

Db

RESULT 29

LOCUS

CK670453 749 bp mRNA linear EST 19-JAN-2005
 DEFINITION UCRCP01_050_H12_T3 Swingle citrumelo nematode-challenged root CDNA

library - UCRCP01 Citrus x paradisi x Poncirus trifoliata CDNA
 clone UCRCP01_050_T3_H12, mRNA sequence.

ACCESSION

CK670453
 VERSION
 KEYWORDS
 EST.

SOURCE

ORGANISM

Citrus x paradisi x Poncirus trifoliata
 Citrus x paradisi x Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosidae; eurosids II; Sapindales; Rutaceae; Citrus x Poncirus.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1 (bases 1 to 749)
 Close, T.J., Roose, M.L., Becker, O., Darow, J., Federici, C.F.,
 Fenton, R.D., Manamaker, S., Choi, Y.G and Kingan, T.
 Development of EST Resources and New Genetic Markers for California
 Citrus - Swingle citrumelo nematode-challenged roots - UCRCP01
 Unpublished (2005)
 Contact: Timothy J. Close
 Department of Botany & Plant Sciences
 University of California
 Riverside, CA 92521-0124, USA
 Tel: 909-787-3318
 Fax: 909-787-4437
 Email: timothy.close@ucr.edu
 Seq primer: T3.
 Location/Qualifiers

FEATURES

source

1..749
 /organism="Citrus x paradisi x Poncirus trifoliata"
 /mol_type="mRNA"
 /cultivar="Swingle"
 /db_xref="taxon:309804"
 /clone="UCRCP01_050_T3_H12"
 /tissue_type="root"
 /dev_stage="seedling"
 /lab_host="E. coli TUC121"
 /clone_lib="Swingle citrumelo nematode-challenged root
 cDNA library - UCRCP01"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI; Site 2: XhoI; Twelve seedlings were
 transplanted from cones to 4 inch pots of sand mix, then
 grown in greenhouses (Darow, Becker lab). Citrus

CC101282
 LOCUS 780 bp DNA linear GSS 16-APR-2003
 DEFINITION CSU-K34.121C22.SP6 CSU-K34 Aedes aegypti genomic clone
 ACCESSION CSU-K34.121C22, genomic survey sequence.
 VERSION CC101282
 KEYWORDS CC101282.1 GI:29969828
 SOURCE GSS.
 ORGANISM Aedes aegypti (yellow fever mosquito)
 Eukaryota; Metazoa; Archipoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Culicinae; Culicini; Aedes; Stegomyia.
 1 (bases 1 to 780)
 Lotius,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
 End sequencing of Aedes aegypti BACs
 Unpublished (2003)
 Other_GSSs: CSU-K34.121C22.T7
 Contact: Brendan Loftus
 Department of Eukaryotic Genomics
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: enta@tigr.org
 Library was provided by Susan Brown and Dennis Knudson at Colorado State University.
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 1..780
 /organism="Aedes aegypti"
 /mol_type="genomic DNA"
 /db_xref="taxon:7159"
 /clone="CSU-K34.121C22"
 /clone_lib="CSU-K34"
 /note="Vector: pBAC3.6; Site 1: EcoRI; Source DNA: Aedes aegypti; Strain unknown [derived from freshly hatched larvae at the Virus Research Centre, Poona, India. Reference: SINGH, K. R. P., 1967 Cell cultures derived from larvae of Aedes albopictus (Skuse) and Aedes aegypti (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC CCL-125"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 9; Length 780;
 Best Local Similarity 91.7%; Pred. No.3.7e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 2 AAAAAAAAAATCGACGACCAATCT 25
 Db 692 AGAAAAAAAAATCACGACCAATCT 715

RESULT 32
 AGS12988/c 801 bp DNA linear GSS 22-DEC-2004
 LOCUS Mus musculus molossinus DNA, clone:MSMg01-417G09.T7, genomic survey
 DEFINITION sequence.
 ACCESSION AGS12988
 VERSION AGS12988.1 GI:48220401
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus (Japanese wild mouse)
 ORGANISM Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Ade,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriaki,K. and Shiroishi,T.
 Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis

JOURNAL
 PUBMED Genome Res. 14 (12), 2439-2447 (2004)
 15574823
 REFERENCE 2 (bases 1 to 801)
 AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenho-chou,Tsukuba-Ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@sc.riken.jp, URL:http://hgp.9ec.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@tc.riken.jp).
 Tsukuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@tc.riken.jp
 PRIMERS
 Sequencing : T7
 LIBRARY
 Vector : pBAC3.6
 R.site 1 : EcoRI
 R.site 2 : EcoRI.

FEATURES
 source
 1..801
 Location/Qualifiers
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMg01-417G09.T7"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 10; Length 801;
 Best Local Similarity 91.7%; Pred. No.3.7e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 AAAAAAAAAATCGACGACCAATC 24
 Db 654 AAAAAAAAAATCGCGACCAAGC 631

RESULT 33
 B0748113/c 811 bp mRNA linear EST 10-OCT-2002
 LOCUS CH3#017 H09T7 Canine heart normalized cDNA library in pbluescript
 DEFINITION Canis familiaris cDNA clone CH3#017_H09 5', mRNA sequence.
 ACCESSION B0748113
 VERSION B0748113
 KEYWORDS B0748113.1 GI:23699874
 SOURCE EST.
 ORGANISM Canis familiaris (dog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 811)
 Yl,Y., Desai,R., Olate,M., Henthorn,P. and George A.J.
 Expressed sequence tags from Canine heart
 Unpublished (2003)
 Other_ESTs: CH3#017_H09T3
 Contact: George AL
 Division of Genetic Medicine
 Vanderbilt University
 529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA
 Tel: 615 936 2660
 Fax: 615 936 2661
 Email: a.george@vanderbilt.edu
 Seq primer: T7: TAATACGACTCACTATAGGG
 High quality sequence start: 42
 High quality sequence stop: 231.
 Location/Qualifiers

FEATURES

Source 1. 811
/organism="Canis familiaris"
/mol_type="mRNA"
/db_xref="taxon:9615"
/clone="CH3#017 H09"
/tissue_type="heart"
/cell_type="heart"
/dev_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"
/clone_1ib="Canine heart normalized cDNA library in pBluescript"
/note="Organ: heart; Vector: pBluescript; Site_1: 5' of vector NotI; Site_2: 3' of vector EcoRI; Tissue source: dog heart (adult, 30 day - 40 day fetal), right and left atria and ventricle. Dog breed - mixed (beagle, German shepherd, pointer, Irish setter). Library construction: oligo-dt primed"

Query Match 83.2%; Score 20.8; DB 5; Length 811;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORIGIN

1 AAAAAAAAAATCGAGCAATC 24
401 AAAAAAAAAACGAGCAATC 378

RESULT 34
CX639167/c

LOCUS CX639167 815 bp mRNA linear EST 18-JAN-2005

DEFINITION UCRPT02-39H09.b Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-39H09-018-1-6.b, mRNA sequence.

ACCESSION CX639167

VERSION CX639167.1 GI:57873996

KEYWORDS EST

SOURCE Poncirus trifoliata

ORGANISM Poncirus trifoliata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eustosids II; Sapindales; Rutaceae; Poncirus.
1 (bases 1 to 815)
Close,T.J., Roose,M.L., Parker,D.R., Federici,C.F., Mandal,J., Fenton,R.D., Wanmaker,S., Landry,B., Hubert,N., Laforest,M., Landry,J. and Ligonde,A.
Development of EST Resources and New Genetic Markers for California Citrus - Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1 Unpublished (2005)
Contact: Timothy J. Close
Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timothy.close@ucr.edu
Seq primer: T7.

JOURNAL COMMENT

TITLE

FEATURES
Source location/Qualifiers
1. 815
/organism="Poncirus trifoliata"
/mol_type="mRNA"
/cultivar="Pomeroy"
/db_xref="taxon:37690"
/clone="UCRPT02-39H09-018-1-6.b"
/tissue_type="root"
/dev_stage="seedling"
/lab_host="E. coli TUC121"
/clone_1ib="Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1"
/note="Vector: lambda uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plant materials were prepared by Federici (Roose lab) with advice from Parker. Seedlings

138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO2 buffering. This solution was replaced on July 6, 2004. Roots from three plants were sampled on June 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stresses in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using TRIzol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanmaker (Close lab) using the HAVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Wanmaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 815;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORIGIN

1 AAAAAAAAAATCGAGCAATC 24
800 AAAAAAAAAACGAGCAATC 777

RESULT 35
AO860645/c

LOCUS AO860645 816 bp DNA linear GSS 03-NOV-1999

DEFINITION nbeb0015B13f CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica cultivar-group) genomic clone nbeb0015B13f, genomic survey sequence.

ACCESSION AO860645

VERSION AO860645.1 GI:6211102

KEYWORDS GSS

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 816)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998)
Contact: Ming RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATCGACTCAGCTTAGGG
Classes: BAC ends
High quality sequence start: 41
High quality sequence stop: 368.
Location/Qualifiers

FEATURES
source

1..816
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbe0015B13f"
/issue_type="leaf"
/lab_host="E. coli DH10B"
/note="CDGI Rice BAC library (ECORI)"
/clone_id="Vector: PBACindigo; Site_1: ECORI; Site_2: ECORI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 816;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAATCGACAAATCT 25
Db 503 AAAAAAAAAACGACAAACT 480

RESULT 36
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

DN957780 825 bp mRNA linear EST 09-MAY-2005
USDA-PP/ARO_13610 Star Ruby grapefruit hot water-treated Flavado
DN957780
DN957780.1 GI:63104514
EST.
Citrus x paradisi
Citrus x paradisi
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

rosids; euroside II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 825)
McCollum, T.G., Maul, P. and Porat, R.
Expressed Sequence Tags (ESTs) from Flavado of Star Ruby grapefruit after hot water treatment
Unpublished (2005)
Contact: McCollum, T.G.
US Horticultural Research Laboratory
USDA, ARS
2001 S. Rock Road, Ft. Pierce, FL 34945, USA
Tel: 561-462-5836
Fax: 561-462-5986
Email: gmcollum@uhrl.ars.usda.gov
Seq primer: T3 primer
Location/Qualifiers

FEATURES
source

1..825
/organism="Citrus x paradisi"
/mol_type="mRNA"
/db_xref="taxon:37656"
/clone="Hw-03_E01"
/issue_type="Flavado"
/dev_stage="mature fruit"
/lab_host="SOLR"
/clone_id="Star Ruby grapefruit hot water-treated Flavado"
/note="Organ: fruit; Vector: pBluescript II SK+; Site_1: ECORI; Site_2: XhoI; Standard library construction protocols from StrataGene cDNA synthesis kit (Cat No. 200401-5) and Uni-ZAP XR vector kit (Cat No. 237211)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 825;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACAAATC 24
Db 52 AAAAAAAAAATCGACAAATC 75

RESULT 37
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CD573838 832 bp mRNA linear EST 12-JUN-2003
UCRPT01_02_E02_T7 Poncirus trifoliata CTV-challenged cDNA library -
UCR Poncirus trifoliata cDNA clone UCRPT01_02_E02, mRNA sequence.
CD573838
CD573838.1 GI:31669740
EST.
Poncirus trifoliata
Poncirus trifoliata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; euroside II; Sapindales; Rutaceae; Poncirus.
1 (bases 1 to 832)
Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D., Manamaker, S., Choi, Y. and Kingan, T.
Development of EST Resources and New Genetic Markers for California Citrus - Poncirus trifoliata CTV-challenged phloem - UCR
Contact: Mikeal Roose
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124, USA
Tel: 9097874137
Fax: 9097874437
Email: mikeal.roose@ucr.edu
Seq primer: T7.
Location/Qualifiers

FEATURES
source

1..832
/organism="Poncirus trifoliata"
/mol_type="mRNA"
/cultivar="Pomeroy Op"
/db_xref="taxon:37690"
/clone="UCRPT01_02_E02"

/tissue_type="Phloem"
/dev_stage="10 - 30 cm shoots"
/lab_host="S. coli TJC121"
/clone_1lb="Poncirus trifoliata CTV-challenged cDNA library - UCR"
/note="vector: lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse at University of California, Riverside. The section was a open-pollinated (very probably selfed) seedling of Poncirus trifoliata cv Pomeroy that was selected as homozygous for the CTV resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate 7514 over 1 year before sampling (CTV infects sweet orange, but not genotypes carrying the CTV resistance gene. Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of Bluescript SK(-) phagemids. All steps to this point were performed in the M. Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the University of California Riverside Genomics Institute, Core Instrumentation Facility, (Choi, Kingan). Chromatogram files were transmitted to UC Riverside (by Choi), then processed at UC Riverside (by Mamamker) using the Harvest pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 812;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACACAATC 24
Db 752 AAAAAAAAAATCGACACAATC 729

RESULT 38
CC916910/c 849 bp DNA linear GSS 08-AUG-2003
LOCUS t097e12ba.f1 TAMBt Bos taurus genomic clone t097e12ba, genomic
DEFINITION survey sequence.

ACCESSION CC916910
VERSION CC916910.1 GI:33545839
KEYWORDS GSS.

SOURCE Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 849)
AUTHORS Lin, S., Najjar, F. Z., Adelson, D., Gill, C. A. and Roe, B. A.
TITLE Bovine BAC End Sequences from Library TAMBt
JOURNAL Unpublished (2003)
CONTACT: Bruce A. Roe
Advanced Center for Genome Technology
University of Oklahoma Department of Chemistry and Biochemistry
620 Parrington Oval, Room 208, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Class: BAC ends
High quality sequence start: 80
High quality sequence stop: 484.

FEATURES

Location/Qualifiers

source

1. 849
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="Angus bull T A M U Shoshone Y6 11519666"
/db_xref="taxon:9913"
/clone_1lb="t097e12ba"
/sex="Male"
/cell_type="Blood"
/clone_1lb="TAMBt"
/note="vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; TAMBt Bovine BAC library (Male) produced by Texas A&M University, Department of Animal Science."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 849;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACACAATC 24
Db 175 AAAAAAAAAATCGACACAAGC 152

RESULT 39

BP695374 851 bp mRNA linear EST 22-DEC-2000
LOCUS 602083260F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4247487 5',
DEFINITION mRNA sequence.

ACCESSION BP695374
VERSION BP695374.1 GI:11980782
KEYWORDS EST.
SOURCE Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo

REFERENCE 1 (bases 1 to 851)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1064 row: h column: 16
High quality sequence stop: 584.

FEATURES

source

1. 851
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_1lb="IMAGE:4247487"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1lb="NIH MGC 81"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match

83.2%; Score 20.8; DB 2; Length 851;

Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAATCGCAGCAAAATCT 25

Db 543 AAAAAAAAAATCGCAGCAAAATCT 566

RESULT 40
DU084516/c 875 bp DNA linear GSS 12-AUG-2005

LOCUS
DEFINITION 280249 Tomato HindIII BAC Library Lycopersicon esculentum genomic clone LE_HBa0171123 3, genomic survey sequence.

ACCESSION
VERSION DU084516 GI:72520657

KEYWORDS

SOURCE
ORGANISM Lycopersicon esculentum (Solanum lycopersicum)

REFERENCE
AUTHORS Muelier,L.A., Buelis,R.M., Wang,Y., Tanksey,S.D., Giovannoni,J.J., Van Eck,J. and Stack,S.

TITLE
JOURNAL BAC end sequencing from three Solanum lycopersicon libraries
COMMENT Other GSSs: 219013

FEATURES

source
Location/Qualifiers
1..875
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/cultivar="Heinz 1706"
/db_xref="taxon:4081"
/clone="LE_HBa0171123"
/lab_host="E. coli"
/clone_lib="Tomato HindIII BAC Library"
/note="Vector: pBelOBAC11; site_1: HindIII"
High quality sequence start: 73
High quality sequence stop: 311.
Location/Qualifiers

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 875;
Best Local Similarity 91.7%; Pred. No. 3.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAGCAAAATC 24

Db 458 AAAAAAAAAATCGCAGCAAAATC 435

RESULT 41
AQ028539/c 296 bp DNA linear GSS 30-JUN-1998

LOCUS
DEFINITION CIT-HSP-2314C7.TR CIT-HSP Homo sapiens genomic clone 2314C7,
genomic survey sequence.

ACCESSION
VERSION AQ028539 GI:3268761

KEYWORDS
SOURCE GSS.

ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 296)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,

TITLE
JOURNAL Golden,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)

COMMENT
JOURNAL Unpublished (1998)
Other GSSs: CIT-HSP-2314C7.TR

REFERENCE
AUTHORS Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

REFERENCE
AUTHORS Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse

FEATURES

source
Location/Qualifiers
1..296
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2314C7"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelOBAC11; site_1: HindIII; site_2:
HindIII"

ORIGIN

Query Match 81.6%; Score 20.4; DB 9; Length 296;
Best Local Similarity 95.5%; Pred. No. 5.4e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGCAGCAAA 22

Db 96 AAAAAAAAAATCGCAGCAAA 75

RESULT 42
CA128042

LOCUS
DEFINITION SCAGL2026F08.g LR2 Saccharum officinarum cDNA clone SCAGL2026F08
5', mRNA sequence.

ACCESSION
VERSION CA128042
CA128042.1 GI:35008521

KEYWORDS
SOURCE EST.

ORGANISM Saccharum officinarum

REFERENCE
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

TITLE
JOURNAL The libraries that made SUCSEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6070, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089

REFERENCE
AUTHORS Email: parreda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcv.unesp.br
Plate: 026 row: F column: 08
Seq primer: T7 Promoter Primer.
Location/Qualifiers

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

source

1.449
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCAGLR2026F08"
/lab_host="DH10B"
/clone_lib="LR2"

/note="Organ: Leaf roll from field grown adult plants (small insert library); Vector: pSPori1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [leaf roll from field grown adult plants (small insert library)]. cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://succest.lad.ic.unimelb.br/public>"

ORIGIN

Query Match 81.6%; Score 20.4; DB 6; Length 449;
Best Local Similarity 95.5%; Pred. No. 5.3e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACACAA 22
|||||
17 AAAAAAAAAATCGACACAA 38

RESULT 43
BH766187 522 bp DNA linear GSS 20-MAR-2002
LOCUS BMBAC340B03T7 PSU Brugia malayi Genomic Bac Library 3 Brugia malayi
DEFINITION genomic, genomic survey sequence.
BH766187
ACCESSION BH766187.1 GI:19563951
VERSION
KEYWORDS GSS.
SOURCE Brugia malayi
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Brugia.
1 (bases 1 to 522)
Whitton,C., Daub,J., Quail,M., Hall,N., Foster,J., Ware,J., Ganatra,M., Slatko,B., Barrell,B. and Blaxter,M.
A genome sequence survey of the filarial nematode Brugia malayi: repeats, gene discovery, and comparative genomics
Mol. Biochem. Parasitol. 137 (2), 215-227 (2004)
15383292
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
Sequenced from the Brugia malayi BAC library constructed by Claire Whitton and Dr Mike Quail. The sequence was generated by The Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICAPB, University of Edinburgh, Edinburgh, UK.
Seq primer: T7 (TAATACGACTCATATAGGG)
Classes: BAC ends.
Location/Qualifiers
1..522
/organism="Brugia malayi"
/mol_type="genomic DNA"
/strain="TRS"
/db_xref="taxon:6279"
/sex="Mixed (male and female)"
/tissue_type="whole parasite"
/dev_stage="microfilaria (L1)"
/clone_lib="Brugia malayi Genomic Bac Library 3"

FEATURES
source

/note="Vector: pBACe3.6; Site 1: BamH I; Brugia malayi genomic DNA was partially cleaved with Sau3A I and size fractionated. 7,392 clones were generated with mean insert size ~48 kbp. The library was constructed by Claire Whitton, Blaxter Nematode Genetics Lab, University of Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, The Sanger Centre, Cambridge, UK."

ORIGIN

Query Match 81.6%; Score 20.4; DB 9; Length 522;
Best Local Similarity 95.5%; Pred. No. 5.2e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACACAA 22
|||||
424 AAAAAAAAAATCGACACAA 445

RESULT 44
BH766171 594 bp DNA linear GSS 20-MAR-2002
LOCUS BMBAC340A06T7 PSU Brugia malayi Genomic Bac Library 3 Brugia malayi
DEFINITION genomic, genomic survey sequence.
BH766171
ACCESSION BH766171.1 GI:19563935
VERSION
KEYWORDS GSS.
SOURCE Brugia malayi
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Brugia.
1 (bases 1 to 594)
Whitton,C., Daub,J., Quail,M., Hall,N., Foster,J., Ware,J., Ganatra,M., Slatko,B., Barrell,B. and Blaxter,M.
A genome sequence survey of the filarial nematode Brugia malayi: repeats, gene discovery, and comparative genomics
Mol. Biochem. Parasitol. 137 (2), 215-227 (2004)
15383292
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
Sequenced from the Brugia malayi BAC library constructed by Claire Whitton and Dr Mike Quail. The sequence was generated by The Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICAPB, University of Edinburgh, Edinburgh, UK.
Seq primer: T7 (TAATACGACTCATATAGGG)
Classes: BAC ends.
Location/Qualifiers
1..594
/organism="Brugia malayi"
/mol_type="genomic DNA"
/strain="TRS"
/db_xref="taxon:6279"
/sex="Mixed (male and female)"
/tissue_type="whole parasite"
/dev_stage="microfilaria (L1)"
/clone_lib="Brugia malayi Genomic Bac Library 3"
/note="Vector: pBACe3.6; Site 1: BamH I; Brugia malayi genomic DNA was partially cleaved with Sau3A I and size fractionated. 7,392 clones were generated with mean insert size ~48 kbp. The library was constructed by Claire Whitton, Blaxter Nematode Genetics Lab, University of Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, The Sanger Centre, Cambridge, UK."

ORIGIN

Query Match 81.6%; Score 20.4; DB 9; Length 594;
Best Local Similarity 95.5%; Pred. No. 5.2e+03;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACAAA 22
 Db 421 AAAAAAAAAATCGACGACAAA 442

RESULT 45
 BH795414/C
 LOCUS 628 bp DNA linear GSS 09-APR-2002
 DEFINITION BMBAC376G03T7_P5U Brugia malayi Genomic Bac Library 3 Brugia malayi
 ACCESSION BH795414
 VERSION BH795414.1 GI:20043746
 KEYWORDS GSS
 SOURCE Brugia malayi
 ORGANISM Brugia malayi
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Brugia.

REFERENCE
 AUTHORS Whitton,C., Daub,J., Quail,M., Hall,N., Foster,J., Ware,J., Ganatra,M., Slatko,B., Barrell,B. and Blaxter,M.
 TITLE A genome sequence survey of the filarial nematode Brugia malayi: repeats, gene discovery, and comparative genomics
 JOURNAL Mol. Biochem. Parasitol. 137 (2), 215-227 (2004)
 PUBMED 15383292
 COMMENT Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk

Sequenced from the Brugia malayi BAC library constructed by Claire Whitton and Dr Mike Quail. The sequence was generated by The Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICARB, University of Edinburgh, Edinburgh, UK
 Seq primer: T7 (TAATACGACTCATATAGG)
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..628
 /organism="Brugia malayi"
 /mol_type="genomic DNA"
 /strain="TBS"
 /db_xref="taxon:6279"
 /sex="Mixed (male and female)"
 /tissue_type="whole parasite"
 /dev_stage="microfilaria (L1)"
 /clone_lib="Brugia malayi Genomic Bac Library 3"
 /note="Vector: pBAC3.6; Site_1: BamH I; Brugia malayi genomic DNA was partially cleaved with Sau3A I and size fractionated. 7,392 clones were generated with mean insert size ~48 kbp. The library was constructed by Claire Whitton, Blaxter Nematode Genetics Lab, University of Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, The Sanger Centre, Cambridge, UK."

ORIGIN
 Query Match 81.6%; Score 20.4; DB 9; Length 628;
 Best Local Similarity 95.5%; Pred. No. 5.2e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACAAA 22
 Db 397 AAAAAAAAAATCGACGACAAA 376

RESULT 46
 BG849593
 LOCUS 635 bp mRNA linear EST 29-MAY-2001
 DEFINITION 1024025H11.x1 C. reinhardtii CC-1690, normalized, lambda Zap II

ACCESSION
 VERSION BG849593
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE
 AUTHORS Grossman,A., Davies,J., Federaple,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants: project phase 2
 JOURNAL Unpublished (2000)
 COMMENT Contact: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chausser@duke.edu.

FEATURES
 source
 Location/Qualifiers
 1..635
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI. This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exs81st (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN
 Query Match 81.6%; Score 20.4; DB 2; Length 635;
 Best Local Similarity 95.5%; Pred. No. 5.2e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGACGACAAA 22
 Db 149 AAAAAAAAAATCGACGACAAA 170

RESULT 47
 A2524641/C
 LOCUS 654 bp DNA linear GSS 07-MAY-2001
 DEFINITION 2349Bc01 Pb MBN #21 Plasmodium berghei genomic 3', genomic survey sequence.

ACCESSION
 VERSION A2524641
 KEYWORDS A2524641.1 GI:13964719
 SOURCE GSS.
 ORGANISM Plasmodium berghei
 Plasmodium berghei
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE
 AUTHORS Carlton,J.M.-R. and Dame,J.B.
 TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
 JOURNAL Parasitol. Today 16 (10), 409 (2000)
 PUBMED 11006469
 COMMENT Contact: Dame JB
 Dept. of Pathobiology, College of Veterinary Medicine
 University of Florida

2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: damej@mail.vetmed.ufl.edu
 Seq primer: M13(-20) forward
 Class: shotgun.

FEATURES

source

Location/Qualifiers
 1..654
 /organism="Plasmodium berghei"
 /mol_type="genomic DNA"
 /strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"
 /db_xref="taxon:5821"
 /dev_stage="asexual blood forms"
 /lab_host="Mus musculus"
 /clone_1lb="Pb MBN #21"
 /note="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site 1: EcoRV; Site 2: EcoRV; Genomic DNA was prepared from asynchronous blood stage forms of the cloned ANKA isolate of P. berghei grown in laboratory Swiss white mice. The DNA was purified from contaminating host DNA by Hoechst Dye 33258-CsCl ultracentrifugation and precipitated. Purified DNA was digested with mung bean nuclease in the presence of 36-38% formamide at 50 C, as described (Vernick, K.D., Imbercki, R.B., and McCutchan, T.F., 1988. Nucleic Acids Research 16:6883-6896). The ends of the digestion fragments were polished using T4 DNA polymerase, and the fragments size selected in the range 500-2000 bp. These were ligated into the EcoRV-cleaved and dephosphorylated pBluescript SK(+) vector. Recombinant plasmids were used to transform E. coli XL10-Gold host cells."

ORIGIN

Query Match 81.6%; Score 20.4; DB 9; Length 654;
 Best Local Similarity 95.5%; Pred. No. 5.2e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACACAAA 22
 |||||
 Db 60 AAAAAAAAAATCGACACAAA 39

RESULT 48

CZ560510

DEFINITION 661 bp DNA linear GSS 17-JUN-2005
 1.P.BM007F03Q.F03.06 WAG Gallus gallus genomic clone WAG-7F3,
 genomic survey sequence.

ACCESSION

CZ560510

KEYWORDS

GSS.

SOURCE

ORGANISM

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

Gallus gallus

ORIGIN

/db_xref="taxon:9031"
 /clone="WAG-7F3"
 /sex="Female"
 /clone_1lb="WAG"
 /note="Vector: pECBAC1; cloning site HINDIII"

Query Match 81.6%; Score 20.4; DB 10; Length 661;
 Best Local Similarity 95.5%; Pred. No. 5.2e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGACACAAA 22
 |||||
 Db 114 AAAAAAAAAATCGACACAAA 135

RESULT 49

CC155322

DEFINITION 805 bp DNA linear GSS 25-APR-2003
 CSU-K34.124F7.SP6 CSU-K34 Aedes aegypti genomic clone
 CSU-K34-124F7, genomic survey sequence.

ACCESSION

CC155322

KEYWORDS

GSS.

SOURCE

ORGANISM

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

Aedes aegypti

```

DEFINITION 602538553f1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4659564 5',
            mRNA sequence.
ACCESSION  BG495917
VERSION    BG495917.1 GI:13457433
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE  1 (bases 1 to 1262)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNLN at:
            http://image.llnl.gov
            Plate: L10CM1455 row: j column: 13
            High quality sequence stop: 88.
FEATURES
  source
    1..1262
    location/Qualifiers
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:4659564"
      /tissue_type="mucoepidermoid carcinoma"
      /lab_host="DH10B (T1 phage-resistant)"
      /clone_lib="NIH_MGC_59"
      /note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site_1:
      SfiI (ggccgctcgcc); Site_2: SfiI (ggccatcgcc);
      Double-stranded cDNA was prepared from cell line RNA. 5'
      and 3' adaptors were used in cloning as follows: 5'
      adaptor sequence: 5'-CACGGCCATTATGCC-3' and 3' adaptor
      sequence: 5'-ATTCTAGAGGCCGAGCGCCGACATG-dt(30)BN-3'
      (where B = A, C, or G and N = A, C, G, or T). Average
      insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
      contained inserts by PCR. This library was enriched for
      full-length clones and was constructed by Clontech
      Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
      library."
ORIGIN
Query Match      81.6%; Score 20.4; DB 2; Length 1262;
Best Local Similarity 95.5%; Pred. No. 5e+03; 1; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAATCGCAGACAAA 22
   |||||
Db 650 AAAAAAAAAATCGCAGACAAA 671

Search completed: December 14, 2005, 07:34:37
Job time : 1760.1 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:43:33 ; Search time 180.2 Seconds
(without alignments)
68.002 Million cell updates/sec

Title: US-10-681-773-2

Perfect score: 25

Sequence: 1 aaaaaaaaaatcgacgaacaatct 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4161359 seqs, 245077644 residues

Total number of hits satisfying chosen parameters: 8322718

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications_NA_New.*

1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3.*
9: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	80.8	182190	7	US-11-121-086-102
2	20.2	80.8	196200	7	US-11-121-086-9
3	20.2	80.8	199321	7	US-11-121-086-10
4	19.8	79.2	2141	6	US-10-750-185-43265
5	19.2	76.8	201	6	US-10-995-561-71043
6	19.2	76.8	928	6	US-10-750-185-35318
7	19.2	76.8	403278	6	US-10-995-561-13421
8	19.2	76.0	157224	7	US-11-112-908-51
9	19.2	76.0	170189	7	US-11-112-908-50
10	18.8	75.2	1112	9	US-11-014-071-1
11	18.8	75.2	131855	7	US-11-112-908-29
12	18.8	75.2	143389	7	US-11-112-908-30
13	18.8	75.2	147700	6	US-10-857-780-3
14	18.8	75.2	150314	7	US-11-112-908-24
15	18.8	75.2	166020	7	US-11-112-908-28
16	18.6	74.4	201	6	US-10-995-561-28667
17	18.6	74.4	201	6	US-10-995-561-31318
18	18.6	74.4	201	6	US-10-995-561-41795
19	18.6	74.4	201	6	US-10-995-561-41797
20	18.6	74.4	201	6	US-10-995-561-42212
21	18.6	74.4	998	6	US-10-750-185-55963
22	18.6	74.4	1010	6	US-10-689-742-71
23	18.6	74.4	1029	6	US-10-750-185-53443

24	18.6	74.4	1051	6	US-10-750-185-38723
25	18.6	74.4	1078	6	US-10-750-185-32659
26	18.6	74.4	1116	6	US-10-750-185-55690
27	18.6	74.4	1756	6	US-10-750-185-31694
28	18.6	74.4	1864	6	US-10-750-185-38225
29	18.6	74.4	2743	6	US-10-821-234-267
30	18.6	74.4	2761	6	US-10-750-185-50437
31	18.6	74.4	2840	6	US-10-750-185-54058
32	18.6	74.4	3156	6	US-10-750-185-46888
33	18.6	74.4	3648	6	US-10-750-185-46658
34	18.6	74.4	3892	6	US-10-750-185-56733
35	18.6	74.4	65931	6	US-10-995-561-13254
36	18.6	74.4	126552	7	US-11-121-086-1
37	18.6	74.4	150314	7	US-11-112-908-24
38	18.6	74.4	150481	7	US-11-112-908-37
39	18.6	74.4	169047	7	US-11-121-086-15
40	18.6	74.4	179892	7	US-11-112-908-39
41	18.6	74.4	191684	7	US-11-121-086-2
42	18.6	74.4	222094	6	US-10-995-561-13244
43	18.6	74.4	268685	6	US-10-933-025-22
44	18.6	74.4	1125000	6	US-10-995-561-13286
45	18.2	72.8	201	6	US-10-995-561-16817
46	18.2	72.8	201	6	US-10-995-561-20188
47	18.2	72.8	201	6	US-10-995-561-28334
48	18.2	72.8	201	6	US-10-995-561-31314
49	18.2	72.8	201	6	US-10-995-561-52582
50	18.2	72.8	201	6	US-10-995-561-52662
51	18.2	72.8	799	6	US-10-750-185-58332
52	18.2	72.8	1020	6	US-10-750-185-25349
53	18.2	72.8	1026	6	US-10-750-185-33249
54	18.2	72.8	1423	6	US-10-750-185-41542
55	18.2	72.8	1674	6	US-10-750-185-26307
56	18.2	72.8	1975	6	US-10-750-185-55045
57	18.2	72.8	3183	6	US-10-750-185-32109
58	18.2	72.8	130472	6	US-10-995-561-13312
59	18.2	72.8	141121	6	US-10-995-561-13262
60	18.2	72.8	170995	7	US-11-121-086-35
61	18.2	72.8	175023	7	US-11-121-086-18
62	18.2	72.8	195235	6	US-10-995-561-13495
63	18.2	72.8	202600	7	US-11-112-908-31
64	18.2	72.8	322094	6	US-10-995-561-13244
65	18.2	72.8	321019	6	US-10-995-561-13204
66	18.2	72.8	380749	6	US-10-995-561-13216
67	17.8	71.2	867	6	US-10-750-185-37322
68	17.8	71.2	937	6	US-10-750-185-53545
69	17.8	71.2	976	6	US-10-750-185-27352
70	17.8	71.2	1121	6	US-10-750-185-54728
71	17.8	71.2	1176	7	US-11-084-624-13
72	17.8	71.2	13928	6	US-10-995-561-13449
73	17.8	71.2	145382	6	US-10-995-561-13272
74	17.8	71.2	398287	6	US-10-995-561-13396
75	17.6	70.4	201	6	US-10-995-561-13572
76	17.6	70.4	201	6	US-10-995-561-13823
77	17.6	70.4	201	6	US-10-995-561-13851
78	17.6	70.4	201	6	US-10-995-561-13851
79	17.6	70.4	600	6	US-10-750-185-40443
80	17.6	70.4	600	6	US-10-750-185-839
81	17.6	70.4	790	6	US-10-750-185-1844
82	17.6	70.4	1273	6	US-10-750-185-40639
83	17.6	70.4	1296	6	US-10-750-185-26600
84	17.6	70.4	1435	6	US-10-750-185-26145
85	17.6	70.4	1453	6	US-10-750-185-31243
86	17.6	70.4	1544	6	US-10-750-185-51622
87	17.6	70.4	1908	6	US-10-750-185-57779
88	17.6	70.4	2126	6	US-10-750-185-57779
89	17.6	70.4	2646	6	US-10-750-185-10064
90	17.6	70.4	2652	6	US-10-750-185-57390
91	17.6	70.4	3712	6	US-10-750-185-61739
92	17.6	70.4	3855	6	US-10-750-185-60547
93	17.6	70.4	4007	6	US-10-750-185-18457
94	17.6	70.4	7500	6	US-10-995-561-13194
95	17.6	70.4	101786	7	US-11-117-187-199
96	17.6	70.4	130472	6	US-10-995-561-13312

Sequence 38723, A
Sequence 32659, A
Sequence 55690, A
Sequence 31694, A
Sequence 38225, A
Sequence 267, App
Sequence 50437, A
Sequence 54058, A
Sequence 46888, A
Sequence 46658, A
Sequence 56733, A
Sequence 13254, A
Sequence 1, Appli
Sequence 24, Appli
Sequence 37, Appli
Sequence 15, Appli
Sequence 39, Appli
Sequence 2, Appli
Sequence 13244, A
Sequence 22, Appli
Sequence 13286, A
Sequence 16817, A
Sequence 20188, A
Sequence 28334, A
Sequence 33134, A
Sequence 52582, A
Sequence 52662, A
Sequence 58332, A
Sequence 25349, A
Sequence 33249, A
Sequence 41542, A
Sequence 26307, A
Sequence 54045, A
Sequence 32109, A
Sequence 13312, A
Sequence 13262, A
Sequence 35, Appli
Sequence 18, Appli
Sequence 13495, A
Sequence 31, Appli
Sequence 13204, A
Sequence 13244, A
Sequence 13216, A
Sequence 37322, A
Sequence 53545, A
Sequence 27352, A
Sequence 54728, A
Sequence 13, Appli
Sequence 13449, A
Sequence 13372, A
Sequence 13396, A
Sequence 13572, A
Sequence 13823, A
Sequence 13851, A
Sequence 40443, A
Sequence 839, App
Sequence 1844, App
Sequence 40639, A
Sequence 34199, A
Sequence 26600, A
Sequence 26145, A
Sequence 31243, A
Sequence 51622, A
Sequence 57779, A
Sequence 57779, A
Sequence 50668, A
Sequence 30864, A
Sequence 57390, A
Sequence 61739, A
Sequence 60547, A
Sequence 18457, A
Sequence 13194, A
Sequence 199, App
Sequence 13312, A

```
97 17.6 70.4 152335 7 US-11-121-086-73 Sequence 73, Appl
98 17.6 70.4 158692 7 US-11-121-086-30 Sequence 30, Appl
99 17.6 70.4 172147 7 US-11-112-908-22 Sequence 22, Appl
100 17.6 70.4 176503 7 US-11-121-086-53 Sequence 53, Appl
101 17.6 70.4 176504 7 US-11-121-086-53 Sequence 53, Appl
102 17.6 70.4 180654 7 US-11-121-086-58 Sequence 58, Appl
103 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
104 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
105 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
106 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
107 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
108 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
109 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
110 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
111 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
112 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
113 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
114 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
115 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
116 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
117 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
118 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
119 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
120 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
121 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
122 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
123 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
124 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
125 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
126 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
127 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
128 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
129 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
130 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
131 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
132 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
133 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
134 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
135 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
136 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
137 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
138 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
139 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
140 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
141 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
142 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
143 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
144 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
145 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
146 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
147 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
148 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
149 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
150 17.6 70.4 184868 7 US-11-121-086-88 Sequence 88, Appl
```

ALIGNMENTS

```
RESULT 1
; Sequence 102, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2004-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; NUMBER OF SEQ ID NOS: 107
```

```
SOFTWARE: PatentIn version 3.3
; SEQ ID NO 102
; LENGTH: 182190
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-102
```

```
Query Match 80.8%; Score 20.2; DB 7; Length 182190;
Best Local Similarity 88.0%; Pred. No. 70;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGACAGCAATCT 25
Db 49368 AAAAAAAAAATCGACAGCAATCT 49392
```

```
RESULT 2
; Sequence 9, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9
; LENGTH: 196200
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-9
```

```
Query Match 80.8%; Score 20.2; DB 7; Length 196200;
Best Local Similarity 88.0%; Pred. No. 70;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGACAGCAATCT 25
Db 67349 AAAAAAAAAATCGACAGCAATCT 67373
```

```
RESULT 3
; Sequence 10, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10
; LENGTH: 199321
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-10
```

```
Query Match 80.8%; Score 20.2; DB 7; Length 199321;
Best Local Similarity 88.0%; Pred. No. 70;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAATCGACAGCAATCT 25
Db 67349 AAAAAAAAAATCGACAGCAATCT 67373
```

Db 190751 AAAAAAAAAATCAGCAAAATCT 190775

RESULT 4

US-10-750-185-43265/c
 ; Sequence 43265, Application US/10750185
 ; Publication NO. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 43265
 ; LENGTH: 2141
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866880135089
 US-10-750-185-43265

Query Match 79.2%; Score 19.8; DB 6; Length 2141;
 Best Local Similarity 91.3%; Pred. No. 52;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAT 23

Db 1777 AAAAAAAAAATCAGCAAAAT 1755

RESULT 5

US-10-995-561-71043/c
 ; Sequence 71043, Application US/10995561
 ; Publication NO. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 71043
 ; LENGTH: 201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-995-561-71043

Query Match 76.8%; Score 19.2; DB 6; Length 201;
 Best Local Similarity 87.5%; Pred. No. 62;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAATC 24

Db 70 AAAAAAAAAATTGAAACAAATC 47

RESULT 6

US-10-750-185-35318
 ; Sequence 35318, Application US/10750185
 ; Publication NO. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KERR, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-2

; CURRENT APPLICATION NUMBER: US/10/750,185

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 35318

; LENGTH: 928

; TYPE: DNA

; ORGANISM: Bovine 19866881454617

US-10-750-185-35318

Query Match 76.8%; Score 19.2; DB 6; Length 928;
 Best Local Similarity 87.5%; Pred. No. 77;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAATC 24

Db 170 AAAAAAAAAAGCAACAAATC 193

RESULT 7

US-10-995-561-13421/c
 ; Sequence 13421, Application US/10995561
 ; Publication NO. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13421
 ; LENGTH: 403278
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1) ... (403278)
 ; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
 US-10-995-561-13421

Query Match 76.8%; Score 19.2; DB 6; Length 403278;
 Best Local Similarity 87.5%; Pred. No. 1,7402;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAATC 24

Db 73644 AAAAAAAAAATTGAAACAAATC 73621

RESULT 8

US-11-112-908-51
 ; Sequence 51, Application US/11112908
 ; Publication NO. US20050260659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Davis, Lisa M.
 ; APPLICANT: Davis, Lisa M.
 ; TITLE OF INVENTION: Breast Cancer Biomarkers
 ; FILE REFERENCE: 04-164-US
 ; CURRENT APPLICATION NUMBER: US/11/112,908
 ; CURRENT FILING DATE: 2005-04-22

```

; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 51
; LENGTH: 157224
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-51

Query Match      76.0%; Score 19; DB 7; Length 157224;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 AAAAAAAAAATCGCAGAC 19
Db      75471 AAAAAAAAAATCGCAGAC 75489

RESULT 9
US-11-112-908-50
; Sequence 50, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 50
; LENGTH: 170189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-50

Query Match      76.0%; Score 19; DB 7; Length 170189;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 AAAAAAAAAATCGCAGAC 19
Db      108231 AAAAAAAAAATCGCAGAC 108249

RESULT 10
US-11-014-071-1
; Sequence 1, Application US/11014071
; Publication No. US20050246796A1
; GENERAL INFORMATION:
; APPLICANT: Cigan, Andrew M.
; APPLICANT: Fox, Timothy W.
; APPLICANT: Hershey, Howard P.
; APPLICANT: Unger, Erica
; APPLICANT: Wu, Yongzhong
; TITLE OF INVENTION: Dominant Gene Suppression Transgenes and
; FILE REFERENCE: Methods of Using Same
```

```

; FILE REFERENCE: 1554
; CURRENT APPLICATION NUMBER: US/11/014,071
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: 60/530,478
; PRIOR FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: 60/591,975
; PRIOR FILING DATE: 2004-07-29
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: Promoter
; LOCATION: (1)...(1112)
; OTHER INFORMATION: P67
US-11-014-071-1

Query Match      75.2%; Score 18.8; DB 9; Length 1112;
Best Local Similarity 90.9%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      1 AAAAAAAAAATCGCAGACAA 22
Db      752 AAAAAAAAAATCGCAGACAA 773

RESULT 11
US-11-112-908-29
; Sequence 29, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 29
; LENGTH: 131855
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-29

Query Match      75.2%; Score 18.8; DB 7; Length 131855;
Best Local Similarity 90.9%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      1 AAAAAAAAAATCGCAGACAA 22
Db      129641 AAAAAAAAAATCGCAGACAA 129662

RESULT 12
US-11-112-908-30
; Sequence 30, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Davis, Lisa M.
; APPLICANT: Harris, Cole
; APPLICANT: Harris, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
```

CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 30
LENGTH: 143389
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-30

Query Match 75.2%; Score 18.8; DB 7; Length 143389;
Best Local Similarity 90.9%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCCGACACAAA 22
Db 97414 AAAAAAAAAATCCGACACAAA 97435

RESULT 13
US-10-857-780-3/c
Sequence 3, Application US/10857780
Publication No. US20050272043A1
GENERAL INFORMATION:
APPLICANT: ROTH, RICHARD B.
APPLICANT: BRAUN, ANDREAS
APPLICANT: KAMMERER, STEFAN M.
APPLICANT: NELSON, MATTHEW ROBERTS
APPLICANT: RENELAND, RICHARD HENRY
APPLICANT: HOTAL-WRIGHTSON, CAROLYN R.
TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
FILE REFERENCE: THEREOF
CURRENT APPLICATION NUMBER: US/10/857,780
CURRENT FILING DATE: 2004-05-28
PRIOR APPLICATION NUMBER: 10/723,681
PRIOR FILING DATE: 2003-11-25
PRIOR APPLICATION NUMBER: 60/490,234
PRIOR FILING DATE: 2003-07-24
PRIOR APPLICATION NUMBER: 60/525,239
PRIOR FILING DATE: 2003-11-25
NUMBER OF SEQ ID NOS: 4962
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 147700
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (51510)..(51510)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc.feature
LOCATION: (51526)..(51526)
OTHER INFORMATION: n is a, c, g, or t
US-10-857-780-3

Query Match 75.2%; Score 18.8; DB 6; Length 147700;
Best Local Similarity 90.9%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCCGACACAAA 22
Db 75954 AAAAAAAAAATCCGACACAAA 75933

RESULT 14
US-11-112-908-24
Sequence 24, Application US/11112908
Publication No. US20050260659A1
GENERAL INFORMATION:
APPLICANT: Harris, Cole
APPLICANT: Davis, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 24
LENGTH: 150314
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-24

Query Match 75.2%; Score 18.8; DB 7; Length 150314;
Best Local Similarity 90.9%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCCGACACAAA 22
Db 11881 AAAAAAAAAATCCGACACAAA 11902

RESULT 15
US-11-112-908-28
Sequence 28, Application US/11112908
Publication No. US20050260659A1
GENERAL INFORMATION:
APPLICANT: Harris, Cole
APPLICANT: Davis, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 28
LENGTH: 166020
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-28

Query Match 75.2%; Score 18.8; DB 7; Length 166020;
Best Local Similarity 90.9%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCCGACACAAA 22
Db 143895 AAAAAAAAAATCCGACACAAA 143916

```
RESULT 16
; US-10-995-561-28667
; Sequence 28667, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28667
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-28667
```

```
Query Match          74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 91 AAAAAAAAAAATCGCAGCAAAATAT 115
```

```
RESULT 17
; US-10-995-561-31318/c
; Sequence 31318, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
```

```
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31318
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-31318
```

```
Query Match          74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 100 AAAAAAAAAAATCAAAACCAAAATCT 76
```

```
RESULT 18
; US-10-995-561-41795
; Sequence 41795, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41795
```

```
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-41795
```

```
Query Match          74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 46 AAAAAAAAAAATCCCAAAATCT 70
```

```
RESULT 19
; US-10-995-561-41797
; Sequence 41797, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41797
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-41797
```

```
Query Match          74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 48 AAAAAAAAAAATCCCAAAATCT 72
```

```
RESULT 20
; US-10-995-561-42212
; Sequence 42212, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42212
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-42212
```

```
Query Match          74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAAAATCGCAGCAAAATCT 25
    |||||
Db 36 AAAAAAAAAAATCCCAAAATCT 60
```

```
RESULT 21
```



```
US-10-750-185-55963/C
; Sequence 55963, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55963
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Bovine 19866881392778
US-10-750-185-55963

Query Match      74.4%; Score 18.6; DB 6; Length 998;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 899 AAAAAAAAAATCGCAGCAATCT 875

RESULT 22
US-10-688-742-71/C
; Sequence 71, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavalie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Meiberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 71
; LENGTH: 1010
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-688-742-71

Query Match      74.4%; Score 18.6; DB 6; Length 1010;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 75 AAAAAAAAAATCGCAGCAATCT 51

RESULT 23
US-10-750-185-53443
; Sequence 53443, Application US/10750185
```

```
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53443
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Bovine 19866880435748
US-10-750-185-53443

Query Match      74.4%; Score 18.6; DB 6; Length 1029;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 80 AAAAAAAAAATCGCAGCAATCT 104

RESULT 24
US-10-750-185-38723
; Sequence 38723, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38723
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: Bovine 19866880982763
US-10-750-185-38723

Query Match      74.4%; Score 18.6; DB 6; Length 1051;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAATCT 25
Db 735 AAAAAAAAAATCGCAGCAATCT 759

RESULT 25
US-10-750-185-32659
; Sequence 32659, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
```

```

; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32659
; LENGTH: 1078
; TYPE: DNA
; ORGANISM: Bovine 19866881014946
US-10-750-185-32659
```

```

Query Match          74.4%; Score 18.6; DB 6; Length 1078;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAAATCGCAGCAATCT 25
Db 843 AAAAAACAATCTCAGCAATCT 867
```

```

RESULT 26
US-10-750-185-55690/c
; Sequence 55690, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55690
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Bovine 19866880853310
US-10-750-185-55690
```

```

Query Match          74.4%; Score 18.6; DB 6; Length 1116;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAAATCGCAGCAATCT 25
Db 971 AAAAAAAAAAGCAGCAATCT 947
```

```

RESULT 27
US-10-750-185-31694
; Sequence 31694, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
```

```

; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31694
; LENGTH: 1756
; TYPE: DNA
; ORGANISM: Bovine 19866881219437
US-10-750-185-31694
```

```

Query Match          74.4%; Score 18.6; DB 6; Length 1756;
Best Local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAAATCGCAGCAATCT 25
Db 411 AAAAAATTAATGCAAAAAATCT 435
```

```

RESULT 28
US-10-750-185-38225/c
; Sequence 38225, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38225
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Bovine 19866881818394
US-10-750-185-38225
```

```

Query Match          74.4%; Score 18.6; DB 6; Length 1864;
Best Local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAAATCGCAGCAATCT 25
Db 85 AAAAAAAAAATCTCAGCAGATCT 61
```

```

RESULT 29
US-10-821-234-267/c
; Sequence 267, Application US/10821234
; Publication No. US2005025114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmant, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
```

```

; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pc_seq_genes Version 1.0
; SEQ ID NO 267
; LENGTH: 2743
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-267

Query Match          74.4% Score 18.6; DB 6; Length 2743;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 47 AAAAAAAAAACTCGAGAGCAATCT 23

RESULT 30
US-10-750-185-50437/c
; Sequence 50437, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50437
; LENGTH: 2761
; TYPE: DNA
; ORGANISM: Bovine 19866880498144
US-10-750-185-50437

Query Match          74.4% Score 18.6; DB 6; Length 2761;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 1076 AAAAAAAAAAAGCAATTTT 1052

RESULT 31
US-10-750-185-54058/c
; Sequence 54058, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```

```

; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54058
; LENGTH: 2840
; TYPE: DNA
; ORGANISM: Bovine 19866880568225
US-10-750-185-54058

Query Match          74.4% Score 18.6; DB 6; Length 2840;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 205 AAAAAAAAAAAGATCAATCT 181

RESULT 32
US-10-750-185-46888/c
; Sequence 46888, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46888
; LENGTH: 3156
; TYPE: DNA
; ORGANISM: Bovine 19866881216971
US-10-750-185-46888

Query Match          74.4% Score 18.6; DB 6; Length 3156;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAAAATCT 25
Db 1823 AAAAAAAAAATTGAGCAATCT 1799

RESULT 33
US-10-750-185-46658
; Sequence 46658, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46658
```

LENGTH: 3648
TYPE: DNA
ORGANISM: Bovine 1986880626667
US-10-750-185-46658

Query Match 74.4%; Score 18.6; DB 6; Length 3648;
Best Local Similarity 84.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAATCT 25
Db 496 AAAAAAAAAATCGAGCAATCT 520

RESULT 34
US-10-750-185-56733/c
Sequence 56733, Application US/10750185
Publication No. US20050260603a1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 56733
LENGTH: 3892
TYPE: DNA
ORGANISM: Bovine 1986880673704
US-10-750-185-56733

Query Match 74.4%; Score 18.6; DB 6; Length 3892;
Best Local Similarity 84.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAATCT 25
Db 382 AAAAAAAAAATCGAGCAATCT 358

RESULT 35
US-10-995-561-13254/c
Sequence 13254, Application US/10995561
Publication No. US20050272054a1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: C1001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13254
LENGTH: 65931
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-13254

Query Match 74.4%; Score 18.6; DB 6; Length 65931;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAATCT 25
Db 33986 AAAAAAAAAATCAACCAATCT 33962

RESULT 36
US-11-121-086-1
Sequence 1, Application US/11121086
Publication No. US20050266459a1
GENERAL INFORMATION:
APPLICANT: POULSEN, TIM S.
APPLICANT: NIELSEN, KIRSTEN V.
TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
FILE REFERENCE: 09138, 6000-00000
CURRENT APPLICATION NUMBER: US/11/121,086
CURRENT FILING DATE: 2005-05-04
PRIOR APPLICATION NUMBER: 60/567,570
PRIOR FILING DATE: 2004-05-04
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn version 3.3
SEQ ID NO 1
LENGTH: 126552
TYPE: DNA
ORGANISM: Homo sapiens
US-11-121-086-1

Query Match 74.4%; Score 18.6; DB 7; Length 126552;
Best Local Similarity 84.0%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAATCT 25
Db 118855 AAAAAAAAAATCTTCAATCT 118879

RESULT 37
US-11-112-908-24/c
Sequence 24, Application US/11112908
Publication No. US20050260659a1
GENERAL INFORMATION:
APPLICANT: HARRIS, Cole
APPLICANT: DAVIS, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 24
LENGTH: 150314
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-24

Query Match 74.4%; Score 18.6; DB 7; Length 150314;
Best Local Similarity 84.0%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGAGCAATCT 25
Db 69115 AAAAAAAAAAAGAAACAAATCT 69091

RESULT 38
US-11-112-908-37

```
Sequence 37, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 37
; LENGTH: 150481
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-37

Query Match          74.4%; Score 18.6; DB 7; Length 150481;
Best Local Similarity 84.0%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 148907 AAAAAAAAAAGCGAAGAAACT 148931

RESULT 39
US-11-121-086-15/C
; Sequence 15, Application US/11121086
; Publication No. US2005026459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 15
; LENGTH: 169047
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-15

Query Match          74.4%; Score 18.6; DB 7; Length 169047;
Best Local Similarity 84.0%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 78437 AAAAAAAAAAGTCACAAAGAAATCT 78413

RESULT 40
US-11-112-908-39
; Sequence 39, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
```

```
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 39
; LENGTH: 179892
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-39

Query Match          74.4%; Score 18.6; DB 7; Length 179892;
Best Local Similarity 84.0%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 70896 AAAAAAAAAAGCGAAGAAACT 70920

RESULT 41
US-11-121-086-2
; Sequence 2, Application US/11121086
; Publication No. US2005026459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 191684
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-2

Query Match          74.4%; Score 18.6; DB 7; Length 191684;
Best Local Similarity 84.0%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 24136 AAAAAAAAAATCGTTTACAAATGT 24160

RESULT 42
US-10-995-561-13244
; Sequence 13244, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13244
; LENGTH: 222094
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13244

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 222094;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 17726 AAAAAAAAAATCGCAAAATAT 17750

RESULT 43
US-10-933-025-22/C
; Sequence 22, Application US/10933025
; Publication No. US20050265987a1
; GENERAL INFORMATION:
; APPLICANT: ROSEN, STEVEN
; APPLICANT: HEMMERICH, STEFAN
; APPLICANT: TOMITA, MEGUMI
; TITLE OF INVENTION: Sulfotransferases and methods of use
; TITLE OF INVENTION: Chetcof
; FILE REFERENCE: UCAL-230CON
; CURRENT APPLICATION NUMBER: US/10/933,025
; CURRENT FILING DATE: 2004-09-01
; PRIOR APPLICATION NUMBER: 10/025,966
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/258,577
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 60/267,831
; PRIOR FILING DATE: 2001-09-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 268685
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(268685)
; OTHER INFORMATION: n = A,T,C or G
US-10-933-025-22

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 268685;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 85755 AAAAAAAAAAGACGACAAACT 85731

RESULT 44
US-10-995-561-13286
; Sequence 13286, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13286
; LENGTH: 1125000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (1)...(1125000)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13286

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 1125000;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAATCT 25
Db 439944 AAAAAAAAAATCCCAAAATCT 439968

RESULT 45
US-10-995-561-16817/C
; Sequence 16817, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16817
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-16817

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 6; Length 201;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAAT 23
Db 92 AAAAAAAAAAGCGAGAAAAAT 70

RESULT 46
US-10-995-561-20188/C
; Sequence 20188, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20188
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-20188

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 6; Length 201;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATCGCAGCAAAAT 23
Db 112 AAAAAAAAAAACCTGACAAAT 90

RESULT 47
US-10-995-561-28334/C
```

```
; Sequence 28334, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28334
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-28334

Query Match          72.8%; Score 18.2; DB 6; Length 201;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACACAAT 23
Db 166 AAAAAAAAAAGCGACACAAT 144

RESULT 48
US-10-995-561-33134
; Sequence 33134, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33134
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-33134

Query Match          72.8%; Score 18.2; DB 6; Length 201;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACACAATCT 25
Db 96 AAAAAAAAAAAGCTCAGACAATTT 120

RESULT 49
US-10-995-561-52582
; Sequence 52582, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52582
; LENGTH: 201
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-995-561-52582

Query Match          72.8%; Score 18.2; DB 6; Length 201;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACACAAT 23
Db 118 AAAAAAAAAAAGCGACACAAT 140

RESULT 50
US-10-995-561-52662
; Sequence 52662, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52662
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-52662

Query Match          72.8%; Score 18.2; DB 6; Length 201;
Best Local Similarity 87.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAATCGACACAAT 23
Db 143 AAAAAAAAAAAGCGACACAAT 165

Search completed: December 14, 2005, 11:40:10
Job time : 188.2 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds

(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-3

Perfect score: 25
Sequence: 1 aaaaaaactaaagcttgatcct 25Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pt:*
9: gb_ro:*
10: gb_scs:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_mtg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	24	96.0	28143	8	AC124305 Homo sapi
C 2	24	96.0	142745	8	AC124311 Homo sapi
C 3	24	96.0	179810	8	AC124317 Homo sapi
C 4	23.4	93.6	147404	8	AC004738 Homo sapi
C 5	22.4	89.6	628	10	BV456584 gtf79604.
C 6	22.4	88.6	724	10	BV491883 S221P6108
C 7	22.4	88.6	88311	14	AL358954 Homo sapi
C 8	22.4	89.6	115335	8	AC093811 Homo sapi
C 9	22.4	89.6	137278	9	AC093811 Homo sapi
C 10	22.4	89.6	196721	8	AC019179 Homo sapi
C 11	21.8	87.2	18525	2	AY821561 Manduca s
C 12	21.8	87.2	84322	15	AP004509 Lotus cor
C 13	21.8	87.2	134089	8	AC112917 Homo sapi
C 14	21.8	87.2	145722	14	AC015503 Homo sapi
C 15	21.8	87.2	160936	14	AC073494 Homo sapi
C 16	21.8	87.2	162223	8	AC145988 Pan trogl
C 17	21.8	87.2	166019	14	AC015648 Homo sapi
C 18	21.8	87.2	187909	14	CR936342 Danio rer

C 19	21.8	87.2	201275	9	AL603662 Mouse DNA
C 20	21.8	87.2	214295	9	AC137947 Mus muscu
C 21	21.8	87.2	217253	14	AC156538 Bos tauru
C 22	21.8	87.2	218875	14	CR931978 Danio rer
C 23	21.8	87.2	227741	9	AC126071 Rattus no
C 24	21.8	87.2	230680	14	AC158901 Mus muscu
C 25	21.8	87.2	236253	14	AC161390 Bos tauru
C 26	21.8	87.2	262198	14	AC106703 Rattus no
C 27	21.4	85.6	1681	15	AK221711 Arabidops
C 28	21.4	85.6	1839	15	AY059899 Arabidops
C 29	21.4	85.6	59793	15	AB025634 Arabidops
C 30	21	84.0	156165	8	AC021453 Homo sapi
C 31	21	84.0	157980	8	AC100852 Homo sapi
C 32	20.8	83.2	654	10	BV267418 S235P6522
C 33	20.8	83.2	4322	6	CQ730691 Sequence
C 34	20.8	83.2	13606	6	AX251316 Sequence
C 35	20.8	83.2	13606	6	AX278803 Sequence
C 36	20.8	83.2	13606	6	AX323700 Sequence
C 37	20.8	83.2	13606	6	AX346713 Sequence
C 38	20.8	83.2	50133	8	AL136095 Human DNA
C 39	20.8	83.2	61569	14	AC105043 Homo sapi
C 40	20.8	83.2	80115	14	AP008123 Homo sapi
C 41	20.8	83.2	104668	8	AC021876 Homo sapi
C 42	20.8	83.2	105199	8	AP001576 Homo sapi
C 43	20.8	83.2	130669	8	AC108667 Homo sapi
C 44	20.8	83.2	133374	14	AC151506 Homo sapi
C 45	20.8	83.2	143252	8	AC105752 Homo sapi
C 46	20.8	83.2	146596	8	AP000497 Homo sapi
C 47	20.8	83.2	151991	8	AP004289 Homo sapi
C 48	20.8	83.2	152224	8	AC144536 Homo sapi
C 49	20.8	83.2	154298	9	AL831773 Mouse DNA
C 50	20.8	83.2	157030	8	AC012537 Homo sapi
C 51	20.8	83.2	157242	9	AC019440 Mus muscu
C 52	20.8	83.2	158105	14	AC023375 Homo sapi
C 53	20.8	83.2	164500	14	AC011201 Homo sapi
C 54	20.8	83.2	167380	8	AP006242 Homo sapi
C 55	20.8	83.2	167780	8	AL353784 Human DNA
C 56	20.8	83.2	170954	14	AC151401 Daeypus n
C 57	20.8	83.2	175336	8	AC023449 Homo sapi
C 58	20.8	83.2	178520	8	AC105015 Homo sapi
C 59	20.8	83.2	180548	8	AL596223 Human DNA
C 60	20.8	83.2	185430	14	AC0252706 Homo sapi
C 61	20.8	83.2	187378	8	AL357150 Human DNA
C 62	20.8	83.2	189975	8	AC109512 Homo sapi
C 63	20.8	83.2	201512	8	AL391833 Human DNA
C 64	20.8	83.2	201641	14	AC148904 Otleleux
C 65	20.8	83.2	202404	14	AC135963 Mus muscu
C 66	20.8	83.2	209278	9	AC158134 Mus muscu
C 67	20.8	83.2	215366	9	AL603836 Mouse DNA
C 68	20.8	83.2	218508	14	AC020639 Homo sapi
C 69	20.8	83.2	246490	14	AC103058 Rattus no
C 70	20.8	83.2	281349	14	CR847855 Homo sapi
C 71	20.4	81.6	110000	1	AE017243_1
C 72	20.4	81.6	110000	1	AE017243_2
C 73	20.4	81.6	110000	1	AE017332_1
C 74	20.4	81.6	112886	8	AL590726 Human DNA
C 75	20.4	81.6	171188	14	AC005025 Homo sapi
C 76	20.4	81.6	173371	14	AC092871 Pan trogl
C 77	20.4	81.6	176350	14	AC092513 Papio anu
C 78	20.4	81.6	176643	14	AC092407 Papio anu
C 79	20.4	81.6	181763	9	AC158661 Mus muscu
C 80	20.4	81.6	191191	8	AC0000119 Homo sapi
C 81	20.4	81.6	199280	14	AC073494 Homo sapi
C 82	20.2	80.8	424	6	CQ472205 Sequence
C 83	20.2	80.8	553	10	BV354105 S230P66RG
C 84	20.2	80.8	608	10	BV327179 S241P6195
C 85	20.2	80.8	617	2	AF148494 Drosophi1
C 86	20.2	80.8	627	6	AR504707 Sequence
C 87	20.2	80.8	627	6	AR519989 Sequence
C 88	20.2	80.8	652	15	GAFF438427 Gloxinia
C 89	20.2	80.8	679	10	BV581519 BV581519
C 90	20.2	80.8	784	10	BV570138 G591P6047
C 91	20.2	80.8	786	10	BV662281 S217P6147

92	20.2	80.8	824	10	BV561998	q1h31a06.
93	20.2	80.8	857	10	BV581086	GS91P6181
94	20.2	80.8	1560	2	AF043416	AF043416 Schistoso
95	20.2	80.8	1694	2	AY343409	AY343409 Schistoso
96	20.2	80.8	2146	2	SMHST1	AF043419 Schistoso
97	20.2	80.8	2154	2	AY343402	AY343402 Schistoso
98	20.2	80.8	2168	2	AY3434010	AY3434010 Schistoso
99	20.2	80.8	2254	2	AY3434015	AY3434015 Schistoso
100	20.2	80.8	2261	2	AF043417	AF043417 Schistoso
101	20.2	80.8	2369	2	AF043418	AF043418 Schistoso
102	20.2	80.8	2785	15	NTUBU14	X77456 N. ladaeum U
103	20.2	80.8	4254	6	COS76746	COS76746 Sequence
104	20.2	80.8	6051	8	HSU52427	U52427 Human RNA p
105	20.2	80.8	6746	6	COS75117	COS75117 Sequence
106	20.2	80.8	8333	6	AX277950	AX277950 Sequence
107	20.2	80.8	8333	6	AX232341	AX232341 Sequence
108	20.2	80.8	8333	6	AX346404	AX346404 Sequence
109	20.2	80.8	9483	6	AX251787	AX251787 Sequence
110	20.2	80.8	9483	6	AX345279	AX345279 Sequence
111	20.2	80.8	9483	6	AX348948	AX348948 Sequence
112	20.2	80.8	11263	14	AC109590	AC109590 Homo sapi
113	20.2	80.8	12698	2	US5372	US5372 Caenorhabdi
114	20.2	80.8	16368	11	AV178049	AV178049 Expressio
115	20.2	80.8	16520	11	AV178048	AV178048 Expressio
116	20.2	80.8	32433	6	AX695521	AX695521 Sequence
117	20.2	80.8	33077	8	AC148474	AC148474 Pan trogl
118	20.2	80.8	34003	8	AY594694	AY594694 Homo sapi
119	20.2	80.8	41991	6	CO869818	CO869818 Sequence
120	20.2	80.8	42105	2	CEM0482	CEM0482
121	20.2	80.8	43374	14	AC092314	AC092314 Homo sapi
122	20.2	80.8	44426	8	AC092314	AC092314 Homo sapi
123	20.2	80.8	46217	14	AC145706	AC145706 Homo sapi
124	20.2	80.8	48008	14	AC165851	AC165851 Bos tauru
125	20.2	80.8	48438	8	HS702119	HS702119 Human DNA
126	20.2	80.8	51096	14	AC165510	AC165510 Bos tauru
127	20.2	80.8	53286	15	AC149490	AC149490 Medicago
128	20.2	80.8	61020	6	AX251546	AX251546 Sequence
129	20.2	80.8	67847	14	AC017241	AC017241 Drosophi
130	20.2	80.8	70453	14	AC101556	AC101556 Mus muscu
131	20.2	80.8	71074	14	AC084721	AC084721 Homo sapi
132	20.2	80.8	72475	8	AL353663	AL353663 Human DNA
133	20.2	80.8	72745	8	AL353663	AL353663 Human DNA
134	20.2	80.8	77332	8	AC087510	AC087510 Homo sapi
135	20.2	80.8	85758	8	AC068645	AC068645 Homo sapi
136	20.2	80.8	86133	14	AC153681	AC153681 Bos tauru
137	20.2	80.8	88899	14	AP008161	AP008161 Locust cor
138	20.2	80.8	89435	14	AC142019	AC142019 Rattus no
139	20.2	80.8	90349	15	AB023036	AB023036 Arabidops
140	20.2	80.8	91393	15	AP006117	AP006117 Locust cor
141	20.2	80.8	95328	8	AC012616	AC012616 Homo sapi
142	20.2	80.8	95567	14	AC091232	AC091232 Homo sapi
143	20.2	80.8	95571	8	AL354480	AL354480 Human DNA
144	20.2	80.8	96153	8	AC092651	AC092651 Homo sapi
145	20.2	80.8	107017	8	AC012510	AC012510 Homo sapi
146	20.2	80.8	110000	9	AE016753_1	Continuation (24 o
147	20.2	80.8	110000	9	AE016753_1	Continuation (2 of
148	20.2	80.8	110000	14	AC110642_0	AC110642 Rattus no
149	20.2	80.8	110000	14	AC134360	AC134360 Rattus no
150	20.2	80.8	110000	14	PFMA13_26	Continuation (27 o

ALIGNMENTS

RESULT 1
AC124305/ 28143 bp DNA linear PRI 06-FEB-2003
LOCUS Homo sapiens chromosome 15, clone RP11-173H16, complete sequence.
DEFINITION AC124305
ACCESSION AC124305.3 GI:28261592
VERSION
KEYWORDS
HTG
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 28143)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-173H16
Unpublished
2 (bases 1 to 28143)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Batra,N., Baeten,V., Bloom,T., Boguslavsky,L., Bouckgalter,B., Brown,A., Camarata,J., Campolongo,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzGerald,M., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karakas,A., Kelle,C., LaRoque,K., Lamazares,R., Landers,T., Lehocaky,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McKernan,K., Meldrum,J., Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,V., Testaye,S., Theodore,J., Topham,K., Travers,M., Trivis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wymn,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 28143)
Birren,B., Nusbaum,C., Lander,E., Abouelell,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Baeten,V., Bloom,T., Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Dearlano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hago,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karakas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrum,J., Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachuka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,V., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wymn,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 28143)
Birren,B., Nusbaum,C., Lander,E., Abouelell,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Baeten,V., Bloom,T., Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Dearlano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hago,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karakas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrum,J., Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M., Talamont, J., Testa, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M. Direct Submission
Submitted (06-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 6, 2003 this sequence version replaced gi:27884912.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27334
Center clone name: 173_H_16

Only the first 28.1 kilobases of this clone are being submitted.
The remainder overlaps accession number AC124997 [WICR project L27341].

FEATURES

SOURCE

Location/Qualifiers
1. 28143
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-173H16"
/clone_lib="RPC1-11 Human Male BAC"
complement(87..203)
repeat_region
/rpt_family="MIR"
570..803
/rpt_family="L1ME2"
1227..1719
/rpt_family="L1ME2"
1720..2030
/rpt_family="AluSx"
2031..2144
/rpt_family="L1ME2"
complement(2150..2547)
/rpt_family="MLT1D"
complement(2558..2694)
/rpt_family="AluUb"
complement(2718..2806)
/rpt_family="MLT1D"
2809..3000
/rpt_family="L1ME2"
complement(3344..3644)
/rpt_family="AluY"
complement(3681..3784)
/rpt_family="MER5A"
3855..3983
/rpt_family="MIR"
3995..4118
/rpt_family="L1ME2"
complement(4112..4753)
/rpt_family="L1MD2"
complement(4754..5040)
/rpt_family="AluSx"
complement(5041..5289)
/rpt_family="L1MD2"
complement(5290..5454)
/rpt_family="FRAM"
complement(5455..6359)
/rpt_family="L1MD2"
complement(6416..7355)
/rpt_family="L1MD2"
complement(7352..8222)
repeat_region

repeat_region /rpt_family="L1MD2"
8309..8511
/rpt_family="L1ME2"
complement(8526..8638)
/rpt_family="AluUb"
8643..8890
/rpt_family="L1ME2"
8891..9182
/rpt_family="AluSx"
9183..9761
/rpt_family="L1ME2"
9762..10057
/rpt_family="AluSx"
10068..10315
/rpt_family="L1ME2"
10318..10374
/rpt_family="MIR"
complement(10626..10933)
/rpt_family="AluSx"
complement(12081..12213)
/rpt_family="FLAM_C"
12527..12562
/rpt_family="L2"
12771..12874
/rpt_family="A-rich"
complement(14178..14485)
/rpt_family="AluSg"
15119..15264
/rpt_family="MIR"
15640..16059
/rpt_family="LTR35"
16060..16105
/rpt_family="TC)n"
16106..16217
/rpt_family="LTR35"
16223..17125
/rpt_family="L1ME3A"
complement(17158..17461)
/rpt_family="AluSx"
complement(17465..17571)
/rpt_family="AluSg/x"
complement(17650..17734)
/rpt_family="BLACKACK"
17745..18059
/rpt_family="L1ME3A"
18065..18154
/rpt_family="MLT1D"

Query Match 96.0%; Score 24; DB 8; Length 28143;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTGATCTT 25
Db 27523 AAAAAAAAACTAAGCTGATCTT 27500

RESULT 2
AC124311 142745 bp DNA linear PRI 08-JAN-2003
LOCUS
DEFINITION Homo sapiens chromosome 15, clone RP11-479F18, complete sequence.
ACCESSION AC124311
VERSION AC124311.7 GI:27545109
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo.
REFERENCE
1 (bases 1 to 142745)
Britten, B., Nusbbaum, C. and Lander, E.
Homo sapiens chromosome 15, clone RP11-479F18
JOURNAL Unpublished

REFERENCE AUTHORS

2. (bases - to 142745)

Birren, B., Linton, L., Nusbaum, C., Landier, E., Ali, A., Allen, N., Anderson, S., Barina, N., Bastien, V., Bloom, T., Bogunilavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, R., Chopeyl, V., Colangelo, M., Collins, S., Collimore, A., Cook, A., Cooke, P., Deatrelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, M., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Gott, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karcas, A., Kells, C., LaRocque, K., Lamaczarski, R., Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McSwan, P., McErmann, K., Meldrum, J., Meneses, L., Mitova, T., Meng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Punnkang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J., Rosettli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, S., Stojanovic, N., Straus, N., Subramanian, A., Talmas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, V., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J. Young, G., Zainoun, J., Zempek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE AUTHORS

3 (bases to 142745)

Birren, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barnes, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, J., Collimore, A., Cook, A., Cooke, P., Dattellaro, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., FitzGerald, M., Gager, D., Galagan, J., Gardner, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, S., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menais, L., Mihova, T., Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Raymond, C., Retta, R., Riese, C., Rogov, P., Roman, J., Roy, A., Schnauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange, S., Thompson, N., Stojanovic, H., Talamas, J., Tesfaye, S., Theodore, J., Tophan, K., Travers, M., Vassiliev, L., Vekrelli, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemel, L., Zimmer, A. and Zody, M.

TITLE
Direct Submission

Submitted (16-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCES

Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 142745)

Barren, B., Nusbacher, C., Lander, E., Ali, A., Allen, N.,
Bairn, N., Bastien, V., Bloom, T., Bogunlavsky, L., Boukhalter, B.,
Camarate, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A.,
Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karakas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., MacLenn, C., Macdonald, P., Major, J.,
Matthews, C., McCarthy, M., Meldrum, J., Menais, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, D., O'Neill, D., Oliver, J.,
Peterson, K., Plunkhant, P., Pierre, N., Raymond, C., Reta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schnauer, S., Schpack, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stoianovic, N., Talamas, T., Testaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wymad, D., Young, G., Zainoun, J., Zambek, L., Zimmer, A., and Zody, M.

WILLIAM, D.: Young, D.
Direct Submission

Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genomics Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (Pages 1 to 142745)
Direct Submission
Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genomics Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (Pages 1 to 142745)

REFERENCE AUTHORS

5 (bases 1 to 142745)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,

TITLE
JOURNAL
COMMENT

Camarate, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fairo, S., Ferreira, P., Fitzgerald, M., Gage, D., Galsano, J., Gargava, S., Gord, S., Graham, L., Grand-Pierre, N., Häfner, N., Hays, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamar, A., Karataas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Ton, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Mensu, L., Mhova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testa, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Vel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (08-JAN-2003) Whitehead Institute/MIT Center for Genomew Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jan 8, 2003 this sequence version replaced gi:24337476.

COMMENT

----- Genome Center
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1B1

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----
Contract number: 137340Center project name: LZ/342
Center phone num: 479 E 18

Center Code Name: 475_F_16

Only the final 142.75 kb of this clone are being submitted. The remainder of the clone is overlapped by accession number AC124997 [MTCR project L27341].

FEATURES

Source

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-479P18"
/clone_lib="RPC1-11 Human Male BAC"
complement(702..1005)
/rpc_family="AlusK"
complement(2557..2848)
/rpc_family="AlusC"
complement(2927..3001)
/rpc_family="MIR"
3058..3081
/rpc_family="AT_rich"
3157..3458
/rpc_family="AluIo"
3661..3780
/rpc_family="AluIo"
4078..4210
/rpc_family="MIR3"
4329..4385
/rpc_family="CA)n"
4597..4623
/rpc_family="CA)n"
4637..4746
/rpc_family="12"
4789..5137
/rpc_family="MT1A1"
complement(5302..5416)
/rpc_family="MIR"
7580..7600
/rpc_family="GC_rich"
7852..7963
/rpc_family="(CCG)n"

```

```

repeat_region      8418 . 8439      /rpt_family="GC-rich"
repeat_region      9209 . 9248      /rpt_family=" (TG) n"
repeat_region      complement(9421 . 9556)
repeat_region      complement(10679 . 10764)
repeat_region      /rpt_family="L2"
repeat_region      complement(10852 . 11162)
repeat_region      /rpt_family="AluY"
repeat_region      11503 . 11541
repeat_region      /rpt_family="MIR"
repeat_region      11545 . 11567
repeat_region      /rpt_family="(TAG)n"
repeat_region      11755 . 12354
repeat_region      /rpt_family="L2"
repeat_region      complement(12348 . 12449)
repeat_region      /rpt_family="L2"
repeat_region      12550 . 12719
repeat_region      /rpt_family="MLT1J"
repeat_region      12720 . 13106
repeat_region      /rpt_family="MLT1A1"
repeat_region      13107 . 13166
repeat_region      /rpt_family="MLT1J"
repeat_region      complement(13271 . 13328)
repeat_region      /rpt_family="L2"
repeat_region      13528 . 13676
repeat_region      /rpt_family="MIR"
repeat_region      13700 . 13816
repeat_region      /rpt_family="MIR"
repeat_region      complement(13823 . 14396)
repeat_region      /rpt_family="L2"
repeat_region      14387 . 14420
repeat_region      /rpt_family="MIR"
repeat_region      14421 . 14452
repeat_region      /rpt_family="L2"
repeat_region      14566 . 14649
repeat_region      /rpt_family="L2"
repeat_region      14930 . 15139
repeat_region      /rpt_family="L2"

```

Query Match 96.0% Score 24; DB 8; Length 142745;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTGATCTT 25
 Db 99820 AAAAAAAAACTAAGCTGATCTT 99843

RESULT 3
 AC124997 179810 bp DNA linear PRI 27-AUG-2002
 LOCUS Homo sapiens chromosome 15, clone RP11-466L14, complete sequence.
 AC124997
 DEFINITION AC124997.4 GI:22507185
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 1 (bases 1 to 179810)
 Birren, B., Nuebaum, C. and Lander, E.
 Homo sapiens chromosome 15, clone RP11-466L14
 Unpublished
 2 (bases 1 to 179810)
 Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barua, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Fitzgerald, W., Gage, D.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Submitted (20-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 179810)
 Birren, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barua, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hago, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karataas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Riese, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J.,
 Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 179810)
 Birren, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barua, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hago, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karataas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Riese, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J.,
 Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 27, 2002 this sequence version replaced gi:22123135.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence submissions@genome.wi.mit.edu
 ----- Project Information


```

repeat_region 110321..110606
                /rpt_family="Alu"
repeat_region 11104..111156
                /rpt_family="MER42"
repeat_region 111716..112012
                /rpt_family="Alu"
repeat_region 112672..112802
                /rpt_family="Alu"
repeat_region 112823..113204
                /rpt_family="MLT1"
repeat_region complement(113370)..113640
                /rpt_family="Alu"
repeat_region 119492..119763
                /rpt_family="MER2"
repeat_region 120060..120270
                /rpt_family="Alu"
repeat_region complement(120965)..121242
                /rpt_family="Alu"
repeat_region complement(121456)..121558
                /rpt_family="MLT1"
repeat_region 124613..124768
                /rpt_family="LTR9"
repeat_region complement(125001)..125710
                /rpt_family="LTR12"
repeat_region 125759..126009
                /rpt_family="LTR9"

```

```

Query Match 93.6%; Score 23.4; DB 8; Length 147404;
Best Local Similarity 96.0%; Pred. No. 95;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 AAAAAAAAACTAAGCTTGATCTT 25
    |||||
Db 87804 AAAAAAAAACTAAGCTTGATCTT 87828

```

```

RESULT 5
BV456584/c 628 bp DNA linear STS 06-APR-2005
LOCUS qtr798e04.b1 Clint Pan troglodytes versus STS genomic, sequence
DEFINITION tagged site.
ACCESSION BV456584
VERSION BV456584.1 GI:62216012
KEYWORDS STS.
SOURCE Pan troglodytes versus
ORGANISM Pan troglodytes versus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Pan.
            1 (bases 1 to 628)
REFERENCE Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
AUTHORS Jaffe,D.B.
            Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
TITLE Human Genome
COMMENT Unpublished (2005)

```

```

JOURNAL
COMMENT Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 628
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human Genome NCBI
            Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes versus), 3 other Pan troglodytes versus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps

```

of unknown origin
(Gen,Unknown Chimp). Common names: Pan troglodytes versus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score >= 20, at least 30% of its base calls must satisfy SNOS(30,25)(single strand NGS, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred score >= 25), and the read must have at least 200 bp SNOS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NGS(30,25) standard was applied to all pairs of overlapping reads to call NGS bases and SNPs. Alignments (between two reads) with less than 100 NGS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (<=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

```

FEATURES
    source
        Location/Qualifiers
            1..628
                /organism="Pan troglodytes versus"
                /mol_type="genomic DNA"
                /sub_species="versus"
                /db_xref="taxon:37012"
                /clone_id="Clint"
                <1..>628

```

```

Query Match 89.6%; Score 22.4; DB 10; Length 628;
Best Local Similarity 95.8%; Pred. No. 1.4e+03;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 2 AAAAAAAAACTAAGCTTGATCTT 25
    |||||
Db 532 AAAAAAAAACTAAGCTTGATCTT 509

```

```

RESULT 6
BV491883 724 bp DNA linear STS 07-APR-2005
LOCUS S221P61080RH1.T0 Yvonne Pan troglodytes troglodytes genomic,
DEFINITION sequence tagged site.
ACCESSION BV491883
VERSION BV491883.1 GI:62335553
KEYWORDS STS.
SOURCE Pan troglodytes troglodytes
ORGANISM Pan troglodytes troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Pan.
            1 (bases 1 to 724)
REFERENCE Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
AUTHORS Jaffe,D.B.
            Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
TITLE Human Genome
COMMENT Unpublished (2005)

```

```

JOURNAL
COMMENT Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted

```


SECRETOR: T6 KNCBPTJCT "NCTBTAJ QUNVHJBAJ BTJH AAB IOT JQCC "NO

Toward a complete human genome sequence
Genome Res. 8 (11): 1097-1108 (1998)

PUBMED 9847074
 REFERENCE 2 (bases 1 to 115335)
 AUTHORS Shah, N. and Hakenson, W.
 TITLE The sequence of Homo sapiens BAC clone RP11-355H11
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 115335)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 115335)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 115335)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 6 (bases 1 to 115335)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 11, 2001 this sequence version replaced gi:115624965.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: MUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0355H11
 Drafting Center: WIBR

 NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

 MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

 SOURCE INFORMATION:
 The RPci-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
 VECTOR: pBACE3.6

 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the right is RP11-240A16, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-355H11; actual end is at base position 80018 of RP11-240A16.

There is an unresolved base at 67647.
 The sequence of AC024662 has been incorporated into AC093811.
 FEATURES
 source
 Location/Qualifiers
 1..115335
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-355H11"
 /clone_11b="RPci-11"
 213..710
 /rpt_family="ERVU"
 repeat_region
 880..1022
 /rpt_family="L2"
 repeat_region
 1315..1568
 /rpt_family="MaLR"
 repeat_region
 1587..1606
 /rpt_family="MIR"
 repeat_region
 2357..2377
 /rpt_family="TTAAA)n"
 repeat_region
 4322..4466
 /rpt_family="MIR"
 repeat_region
 4487..4630
 /rpt_family="MaLR"
 repeat_region
 5402..5425
 /rpt_family="TTTG)n"
 repeat_region
 5460..5487
 /rpt_family="AT_rich"
 repeat_region
 5514..5563
 /rpt_family="TA)n"
 repeat_region
 7007..7144
 /rpt_family="MIR"
 repeat_region
 7156..7254
 /rpt_family="MIR"
 repeat_region
 7764..8243
 /rpt_family="L2"
 repeat_region
 7908..7929
 /rpt_family="AT_rich"
 repeat_region
 8263..8637
 /rpt_family="MaLR"
 repeat_region
 11252..11289
 /rpt_family="(T)n"
 repeat_region
 11265..11553
 /rpt_family="Alu"
 repeat_region
 12887..12909
 /rpt_family="(TG)n"
 repeat_region
 13341..13800
 /rpt_family="BRVK"
 repeat_region
 14357..15693
 /rpt_family="L1"
 repeat_region
 15694..15783
 /rpt_family="L1"
 repeat_region
 15784..16961
 /rpt_family="L1"
 repeat_region
 16941..16962
 /rpt_family="AT_rich"
 repeat_region
 16962..17261
 /rpt_family="MER2_type"
 repeat_region
 17092..17121
 /rpt_family="AT_rich"
 repeat_region
 17893..18071
 /note="match to EST AV690508 (NID:gi0292371)"
 repeat_region
 18617..18644
 /rpt_family="AT_rich"
 repeat_region
 19215..19261
 /rpt_family="MIR"
 repeat_region
 20566..20648
 /rpt_family="(TA)n"
 repeat_region
 20650..20670
 /rpt_family="AT_rich"
 repeat_region
 21827..21909

```

repeat_region      /rpt_family="ERV1"
22535..22735
/rpt_family="L2"
repeat_region      /rpt_family="Alu"
22890..23184
/rpt_family="Alu"
repeat_region      /rpt_family="AT_rich"
23767..23802
/rpt_family="AT_rich"
repeat_region      /rpt_family="Alu"
23903..24193
/rpt_family="Alu"
repeat_region      /rpt_family="AT_rich"
24199..24220
/rpt_family="AT_rich"
repeat_region      /rpt_family="L2"
24349..24535
/rpt_family="L2"
repeat_region      /rpt_family="ERV1"
25337..25460
/rpt_family="ERV1"
repeat_region      /rpt_family="TATATG"
25346..25483
/rpt_family="TATATG"
repeat_region      /rpt_family="MERL_type"
25782..25891
/rpt_family="MERL_type"
repeat_region      /rpt_family="MaLR"
25935..26338
/rpt_family="MaLR"
repeat_region      /rpt_family="T"
26407..26440
/rpt_family="T"
repeat_region      /rpt_family="T"
26411..26704
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
26709..26758
/rpt_family="Alu"
repeat_region      /rpt_family="Mariner"
27106..27147
/rpt_family="Mariner"
repeat_region      /rpt_family="TTTTA"
27135..27269
/rpt_family="TTTTA"
repeat_region      /rpt_family="Alu"
27471..27546
/rpt_family="Alu"
repeat_region      /rpt_family="MIR"
27540..27576
/rpt_family="MIR"
repeat_region      /rpt_family="CAT"
27980..28004
/rpt_family="CAT"
repeat_region      /rpt_family="AT_rich"
27992..28212
/rpt_family="AT_rich"
repeat_region      /rpt_family="L1"
28213..28517
/rpt_family="L1"

```

```

Query Match      89.6% Score 22.4; DB 8; Length 115335;
Best Local Similarity 95.8%; Pred. No. 2.2e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAACTTAAGCTTGATCT 24
Db 113426 AAAAAAAAACTTAAGCTTGATCT 113449

```

```

RESULT 9
AC104103      137278 bp      DNA      linear      ROD 27-NOV-2003
LOCUS      AC104103
DEFINITION      Mus musculus BAC clone RP24-37816 from chromosome 17, complete
sequence.
ACCESSION      AC104103
VERSION      AC104103.5
KEYWORDS      GI:33457243
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE      Nguyen, C. and Bielicki, L.
TITLE      The sequence of Mus musculus BAC clone RP24-37816
JOURNAL      Unpublished (2001)
AUTHORS      Wilson, R.
REFERENCE      2 (bases 1 to 137278)
TITLE      Sequencing of Mus musculus
JOURNAL      Unpublished (2001)
AUTHORS      3 (bases 1 to 137278)
REFERENCE      McPherson, J.D. and Waterston, R.H.

```

```

TITLE      Direct Submission
JOURNAL      Submitted (03-DEC-2001) Genome Sequencing Center, 4444 Forest Park
REFERENCE      Parkway, St. Louis, MO 63108, USA
AUTHORS      4 (bases 1 to 137278)
TITLE      Direct Submission
JOURNAL      Submitted (15-MAY-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE      Parkway, St. Louis, MO 63108, USA
AUTHORS      5 (bases 1 to 137278)
TITLE      Direct Submission
JOURNAL      Submitted (06-AUG-2003) Genome Sequencing Center, 4444 Forest Park
REFERENCE      Parkway, St. Louis, MO 63108, USA
AUTHORS      6 (bases 1 to 137278)
TITLE      Direct Submission
JOURNAL      Submitted (27-NOV-2003) Department of Genetics, Washington
REFERENCE      University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS      On Aug 6, 2003 this sequence version replaced gi:30725975.
JOURNAL      -----
COMMENT      -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
Summary Statistics
Center project name: M_BB0378106

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RP24-37816 BAC library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC122502.

```

FEATURES
source
1..137278
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/map="17"
/clone="RP24-37816"
/clone_id="RP24-37816"
repeat_region      235..318
/rpt_family="L1"
repeat_region      592..1235
/rpt_family="L1"
repeat_region      1383..1828
/rpt_family="L1"
repeat_region      4312..4662
/rpt_family="MaLR"

```

```

repeat_region      4689. .4841
/rpc_family="MIR"
repeat_region      5308. .5597
/rpc_family="B4"
repeat_region      6999. .7142
/rpc_family="B4"
repeat_region      7230. .7945
/rpc_family="L1"
repeat_region      9588. .9705
/rpc_family="B4"
repeat_region      10600. .10830
/rpc_family="MALR"
repeat_region      10944. .10998
/rpc_family="L2"
repeat_region      11030. .11079
/rpc_family="ERV1"
repeat_region      11049. .11095
/rpc_family="ERV1"
repeat_region      11288. .11332
/rpc_family="MALR"
repeat_region      11333. .11654
/rpc_family="MALR"
repeat_region      11655. .11861
/rpc_family="MALR"
repeat_region      11970. .12107
/rpc_family="CR1"
repeat_region      15634. .16165
/rpc_family="L1"
repeat_region      16295. .16412
/rpc_family="MALR"
repeat_region      16768. .16855
/rpc_family="CR1"
repeat_region      17520. .17605
/rpc_family="L2"
repeat_region      17628. .17809
/rpc_family="B4"
repeat_region      18051. .19106
/rpc_family="L1"
repeat_region      19360. .19460
/rpc_family="Alu"
repeat_region      19488. .19605
/rpc_family="MER1_type"
repeat_region      19606. .19750
/rpc_family="Alu"
repeat_region      19751. .19859
/rpc_family="MER1_type"
repeat_region      20055. .20414
/rpc_family="L1"
repeat_region      20637. .20898
/rpc_family="B4"
repeat_region      21414. .21539
/rpc_family="L1"
repeat_region      21628. .22512
/rpc_family="L1"
repeat_region      22699. .23376
/rpc_family="L1"
repeat_region      23854. .23981
/rpc_family="ERVK"
repeat_region      24036. .24312
/rpc_family="L1"
repeat_region      24507. .24612
/rpc_family="L1"
repeat_region      24617. .24974
/rpc_family="MALR"
repeat_region      24975. .25531
/rpc_family="L1"
repeat_region      25576. .25752
/rpc_family="ERVK"
repeat_region      26150. .26448
/rpc_family="L1"
repeat_region      26464. .26564
/rpc_family="BCL_MM"
repeat_region      27043. .27324

```

```

/rpc_family="MALR"
repeat_region      29522. .29602
/rpc_family="MIR"
repeat_region      29650. .29737
/rpc_family="MIR"
repeat_region      31264. .32451
/rpc_family="L2"
repeat_region      33911. .34064
/rpc_family="MALR"
repeat_region      34065. .34137
/rpc_family="ID"
repeat_region      34138. .34405
/rpc_family="MALR"
repeat_region      34408. .34545
/rpc_family="MALR"
repeat_region      35069. .35157
/rpc_family="L1"
repeat_region      35158. .35337
/rpc_family="B4"
repeat_region      35584. .37021
/rpc_family="L1"
repeat_region      38187. .38389
/rpc_family="MER1_type"
repeat_region      38615. .38847
/rpc_family="MALR"
repeat_region      39437. .39751
/rpc_family="L1"
repeat_region      40806. .40995
/rpc_family="B4"
repeat_region      41217. .41349
/rpc_family="B4"
repeat_region      41866. .41909
/rpc_family="MIR"

```

Query Match 89.6%; Score 22.4; DB 9; Length 137278;
 Best Local Similarity 95.8%; Pred. No. 2.1e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCT 24
 Db 84745 AAAAAAAAACTAAGCTGATCT 84768

```

RESULT 10
AC019179      AC019179      196721 bp      DNA      linear      PRI 09-MAY-2001
LOCUS      Homo sapiens BAC clone RP11-240A16 from 4, complete sequence.
DEFINITION      AC019179
ACCESSION      AC019179
VERSION      AC019179.4      GI:11120947
KEYWORDS      HMG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE      1 (bases 1 to 196721)
AUTHORS      Sulston, J.E. and Waterston, R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
PUBMED      9847074
REFERENCE      2 (bases 1 to 196721)
AUTHORS      Harkins, R., Maupin, R., Gregory, S., Coblitz, B. and Fleming, A.
TITLE      The sequence of Homo sapiens BAC clone RP11-240A16
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 196721)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      4 (bases 1 to 196721)
AUTHORS      Waterston, R.
TITLE      Direct Submission

```

JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 156721)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Nov 8, 2000 this sequence version replaced gi:7630907.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watsn.wustl.edu
----- Summary Statistics
Center project name: H_NH0240A16

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPI1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: PBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RPI1-240A16; actual end is at base position 156721 of RPI1-240A16.

The sequence H NH0240A16 from base position 157677 to 158503 contains a tandem repeat. The assembly is consistent with digest information about the sequence fidelity cannot be guaranteed.

FEATURES
source
1..156721
location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RPI1-240A16"
/clone_1fb="RPI1-11"
repeat_region
/rpt_family="ERV"
3448..3840
repeat_region
/rpt_family="ERVK"
4103..4596
repeat_region
/rpt_family="ERVK"
5164..5231
repeat_region
/rpt_family="MIR"
5616..5734
repeat_region
/rpt_family="MIR"
5861..5911
repeat_region
/rpt_family="Mariner"
6619..6749

/rpt_family="MALR"
repeat_region
6727..6776
/rpt_family="MALR"
repeat_region
8458..8684
/rpt_family="Alu"
repeat_region
8685..8843
/rpt_family="Alu"
repeat_region
10345..10912
/rpt_family="MALR"
repeat_region
10952..11592
/rpt_family="L2"
repeat_region
14824..15153
/rpt_family="L1"
repeat_region
16195..16619
/rpt_family="ERVL"
repeat_region
16649..16822
/rpt_family="MALR"
repeat_region
17787..17850
/rpt_family="Alu"
repeat_region
18581..18773
/rpt_family="MER1_type"
repeat_region
19637..19697
/rpt_family="L2"
repeat_region
19798..19971
/rpt_family="Alu"
repeat_region
19991..20128
/rpt_family="Alu"
repeat_region
22389..22580
/rpt_family="MIR"
repeat_region
23974..24287
/rpt_family="Alu"
repeat_region
25165..25399
/rpt_family="MIR"
repeat_region
25434..25719
/rpt_family="Alu"
repeat_region
26014..26157
/rpt_family="L2"
repeat_region
28448..28553
/rpt_family="MIR"
repeat_region
29093..29243
/rpt_family="Alu"
repeat_region
29715..30240
/rpt_family="MALR"
repeat_region
30717..30823
/rpt_family="L1"
repeat_region
30859..30929
/rpt_family="L1"
repeat_region
30935..31516
/rpt_family="ERV1"
repeat_region
31502..31764
/rpt_family="ERV1"
repeat_region
31871..31981
/rpt_family="L1"
repeat_region
31982..32494
/rpt_family="ERVK"
repeat_region
32495..32523
/rpt_family="L1"
repeat_region
32579..32646
/rpt_family="MALR"
repeat_region
32775..32922
/rpt_family="ERVL"
repeat_region
33108..33153
/rpt_family="ERVV"
repeat_region
33187..33486
/rpt_family="Alu"
repeat_region
33588..33696
/rpt_family="MALR"
repeat_region
34154..34461
/rpt_family="L2"
repeat_region
34462..35265
/rpt_family="ERV1"
repeat_region
35266..35432
/rpt_family="L2"

```
repeat_region      35517..35685
                    /rpt_family="MALR"
repeat_region      35868..35939
                    /rpt_family="MIR"
repeat_region      38314..38622
                    /rpt_family="L1"
repeat_region      38628..38765
                    /rpt_family="L1"
repeat_region      38872..39419
                    /rpt_family="L1"
repeat_region      41647..41946
                    /rpt_family="Alu"
repeat_region      42051..42251
                    /rpt_family="L1"
repeat_region      42269..42886
                    /rpt_family="ERV1"
repeat_region      42887..43631
                    /rpt_family="L1"
repeat_region      43803..44108
                    /rpt_family="Alu"
repeat_region      47859..47918
                    /rpt_family="ERV1"
repeat_region      47963..48092
                    /rpt_family="ERV1"
repeat_region      48723..49085
                    /rpt_family="ERV1"
```

Query Match 89.6%; Score 22.4; DB 8; Length 196721;
Best Local Similarity 95.8%; Pred. No. 1.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 24
Db 91 AAAAAAAAACTAAGCTTGATCT 114

RESULT 11
AY821561 18525 bp DNA linear INV 01-MAY-2005
LOCUS Manduca sexta chitin synthase 2 gene, complete cds.
DEFINITION
ACCESSION AY821561
VERSION AY821561.1 GI:60459511

SOURCE
ORGANISM Manduca sexta (tobacco hornworm)

REFERENCE
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Sphingidae; Sphingidae; Sphinginae; Manduca.

1 (bases 1 to 18525)
Hogenkamp, D.G., Arakane, Y., Zimoch, L., Merzendorfer, H.,
Kramer, K.J., Beeman, R.W., Kanost, M.R., Specht, C.A. and
Mutchukriehnan, S.

TITLE Chitin synthase genes in *Manduca sexta*: characterization of a
gut-specific transcript and differential tissue expression of
alternately spliced mRNAs during development
JOURNAL Insect Biochem. Mol. Biol. 35 (6), 529-540 (2005)

PUBMED 15857759
AUTHORS 2 (bases 1 to 18525)
Hogenkamp, D.G., Arakane, Y., Zimoch, L., Merzendorfer, H.,
Kramer, K.J., Beeman, R.W., Kanost, M.R., Specht, C.A. and
Mutchukriehnan, S.

TITLE Direct Submission
JOURNAL Submitted (08-NOV-2004) Biochemistry, Kansas State University, 104
Willard Hall, Manhattan, KS 66506, USA

FEATURES
SOURCE Location/Qualifiers
1..18525

ORIGIN
/organism="Manduca sexta"
/db_type="genomic DNA"
/db_xref="taxon:7130"
join(1..166,448..562,3845..4089,4323..4489,4699..5029,
5228..5474,5706..6015,7146..7366,7867..7952,8189..8340,
9380..9642,9953..10163,10355..10555,11115..11341,
11967..12134,12374..12679,12960..13189,13563..13689,

CDS

```
14169..14321,14879..15055,15715..15882,17379..17610,  
17770..17936,18340..18525)  
/product="chitin synthase 2"  
join(465..562,3845..4089,4323..4489,4699..5029,5228..5424,  
5706..6015,7146..7366,7867..7952,8189..8340,9380..9642,  
9953..10163,10355..10555,11115..11341,11967..12134,  
12374..12679,12960..13189,13563..13689,14169..14321,  
14879..15055,15715..15882,17379..17610,17770..17936,  
18340..18477)  
/BC_number="2.4.1.16"  
/note="MeChS2"
```

```
/codon_start=1  
/product="chitin synthase 2"  
/protein_id="AAK20092.1"  
/db_xref="GI:60459512"
```

```
/translation="MAATTPGFKLADSDSDTETVPLDGDDEIDORTAQETKGN  
LFEIPVKEGSGMATNMTIETSVKIKVLAIVLFCVAGSAVIAGTLLFTISOK  
KDRQITHCNRLADDOPIVHSLBEIRITLMALIVGVLGVLRSVRCFKTA  
KKPTKQFIAPITETIOAIGIAVLIIPELDVAGVLMATCAIPALINIFRD  
RMSKRSIKILIDVLAISAQTAFAVWPLMERTVMTITVACVVLGEMENFDVY  
NKSIVFVLOELNDLNRKYTORVLSVKIIVFACILISLMQNDNPTFTAS  
KAFGEQYVNEVLIVRDEDTIGDYTGFIPELDAIWTSAIWLIOVAAYPCFS  
GFACCKILIONFETLALVGPVAILLAFCGMRADPCAFRTIPDLFYRIPV  
YPLREYVGHMAWMLMLISQAMIVHTWQPCRERLSATDKLPKPKMYGILIDSL  
LNRITDLDNDQVEDIKGIGDSSVSGDLAIYDIPRPSITRPOVCATMHTNEE  
MIEFKSIFRLDEQSKRRVQAQKTYGIVDPDYBELCHIMDDAFEISDAEDSVN  
RFVCLADVDEASEVHLTVNRLEPPEKTYPTPGKLIWMPGKNLICHKDKSKI  
RHRKRSQVVMYVFLGRLMDLEISVDRKEVLAENTYLLADGDIDFKSAVTLVD  
LMKDKNLGAACGRIFPVSGFMAWYMPYEGHMLQKTEHMGVLSGPGFSIF  
RGALMDNDVWKYTLFNSNARHVOYDOEDRMLCTLLORGVRVYSAASDGYCF  
PERDEPFNRNRVPESTMAIFDLASDRVQVONISTVIVYVCMIMGTILCP  
GTIFLMATGAINATGSMNHALFINVLPVITVLVCMTKSETOLMLANLITCFV  
VMFVIVSLVLOISQDGLAPSSMFTATGTFPVNALHPQELICLLYSIYITIP  
SMYMLLIYELCNMNNVSGTREVAKQKEMEMDKAAEBAKKNOSIMKPEK  
SDETSGLSEFVAGLFRMCCTNPDKHDLHLIOLNSIEKIKRSALGAESEPA  
OQOTRRSSSLGRDLSLATMPEVADSLSGDIPREEDDILINPVVADPILQGEVPE  
LTTAIEFPWKDILIDVLRPIDENKEBERIKTDLKLRDTPVAFANLPLVITL  
LOPNODOLHKPFQGDVALSYDKERNVAVBQFLMLETGSLFVFGVMLIOV  
AMLSHRSYITLITLSTELHWYFSRRDDNSDLERKVEIARELQKNTDLDGR  
AVSTNDVSRKTLHNEKARDTGSVNDLANKRRLITLQSGDPNVISLSSIGDE  
VTRRATIRALKTRDLSLAEKRSQLOAADATGYVNLSTAVNDSGRASASTAVI  
NKGYEPAFDSDDEPPRPSYVRFRENTY"
```

ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 18525;
Best Local Similarity 92.0%; Pred. No. 6.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCT 25
Db 11025 AAAAAAAAACTAAGCTTGATCT 11001

RESULT 12
AP004509/c 84322 bp DNA linear PLN 22-JUL-2003
LOCUS Locust corniculatus var. japonicus genomic DNA, Chromosome 6,
clone: Lj16K17, TM0037a, complete sequence.

ACCESSION AP004509
VERSION AP004509.1 GI:17736876
KEYWORDS HTG.
SOURCE Locust corniculatus var. japonicus (locust japonicus)
ORGANISM Locust corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.

REFERENCE
AUTHORS Sato, S., Kaneko, T., Nakamura, Y., Aamizu, E., Kato, T. and Tabata, S.
TITLE Structural Analysis of a Locust japonicus Genome. I. Sequence
Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb
Regions of the Genome
JOURNAL Unpublished

REFERENCE	2 (bases 1 to 84322)
AUTHORS	Nakamura, Y.
TITLE	
JOURNAL	
FEATURES	Direct Submission Submitted (13-DEC-2001) Yasukazu Nakamura, Kazuo DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazuo.or.jp, URL:http://www.kazuo.or.jp, Tel:01-438-52-3935, Fax:01-438-52-3934)
SOURCE	Location/Qualifiers 1..84322 /organism="Lotus corniculatus var. japonicus" /mol_type="genomic DNA" /variety="japonicus" /db_xref="taxon:34305" /chromosome="6" /clone="LJ16K17" /clone_id="LJT library" /note="TM0037a, a part of TAC clone:TM0037 synonym: Lotus japonicus"
ORIGIN	
Query Match	87.2%; Score 21.8; DB 15; Length 84322;
Best Local Similarity	92.0%; Pred. No. 4e+02;
Matches	23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY	1 AAAAAAAAACTAAAGCTTGATCCTT 25
Db	24440 AAAAAAAAAACGAGACTTGATCCTT 24416
LOCUS	
DEFINITION	AC112917 Homo sapiens X BAC RP11-46C18 bp DNA linear PRI 24-SEP-2002
ACCESION	AC112917 Homo sapiens X BAC RP11-46C18 (Rosewell Park Cancer Institute Human BAC Library) complete sequence.
VERSION	AC112917
KEYWORDS	AC112917.3 GI:19807697
SOURCE	HTC.
ORGANISM	Homo sapiens (human) Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 134089) Munzy,D.M., Adam,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alabrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foeter,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homel,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jlay.Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovac,K., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu W., Loussaged,H., Lozardo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Nagai,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Maunhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,G., Orangunye,N., Pace,A., Payton,B., Ogutu,M., Okmonu,G., Otagunye,N., Oyiedo R., Pacey,A., Payton,B., Peethy,J., Perez,U., Peters,L., Pickens,R., Primmis,E., Pu,L.U.,

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

2 (bases 1 to 134089)
Worley,K.C.
Direct Submission
Submitted (25-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 134089)
Worley,K.C.
Direct Submission
Submitted (06-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 134089)
Worley,K.C.
Direct Submission
Submitted (29-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 134089)
Worley,K.C.
Direct Submission
Submitted (02-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 134089)
Worley,K.C.
Direct Submission
Submitted (24-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Mar 29, 2002 this sequence version replaced gi:19172732.

COMMENT
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

repeat_region

```

_region_/rptc_family="MER6"
_region_/complement(11533, 11841)
_region_/rptc_family="AluSx"
_region_11842, .11934
_region_/rptc_family="MER6"
_region_11938, .11965
_region_/rptc_family="(TG)n"
_region_11992, .12156
_region_/rptc_family="(TA)n"
_region_12232, .12258
_region_/rptc_family="(CA)n"
_region_12265, .12432
_region_/rptc_family="(TATATG)n"

ch87.2%; Score 21.8; DB 8; Length 134089;
13Similarity 92.0%; Pred. No. 3,4e+02;
23Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1AAAAAAAAAACTAAGCTTGATCTT 25
17AGAAAAAAATTTAAAGCTTGATCTT 4041

AC015503145722 bp DNA linear HTG 07-DEC-2000
Homo sapiens clone RP11-21H2, WORKING DRAFT SEQUENCE, 15 unordered
pieces.
AC015503AC015503.4 GI:11597075
HTG: HTGS PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1(bases 1 to 145722)
Birren,B., Linton,L., Nuebaum,C. and Lander,E.
Homo sapiens, clone RP11-21H2
Unpublished
2(bases 1 to 145722)
Birren,B., Linton,L., Nuebaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barina,N., Becker,T.Y., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castet,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deatellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferris,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galsagan,J., Gardyana,S., Grant,G., Hagos,B., Heatford,A., Horton,J.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karstad,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McGowan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tetzlaff,S., Tjirell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submissions
Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 7, 2000 this sequence version replaced gi:10334917.
All repeats were identified using RepeatMasker:
Smith,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L4014
Center clone name: 21.H.2
Summary Statistics
Sequencing vector: M13; M77815; 5% of reads
Sequencing vector: Plasmid; n/a; 95% of reads

```



```

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 138497 bases at least Q40
Consensus quality: 142273 bases at least Q30
Consensus quality: 143675 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 144322; sum-of-contigs
Quality coverage: 5.6 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 base.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
* 1 10503: contig of 10503 bp in length
* 10504 10503: gap of 100 bp
* 10604 12771: contig of 2168 bp in length
* 12772 12871: gap of 100 bp
* 12872 14989: contig of 2118 bp in length
* 14990 15089: gap of 100 bp
* 15090 16911: contig of 1822 bp in length
* 16912 17011: gap of 100 bp
* 17012 19684: contig of 2673 bp in length
* 19685 19784: gap of 100 bp
* 19785 22088: contig of 2304 bp in length
* 22089 22188: gap of 100 bp
* 22189 25445: contig of 3257 bp in length
* 25446 25545: gap of 100 bp
* 25546 30628: contig of 5083 bp in length
* 30629 30728: gap of 100 bp
* 30729 37323: contig of 6595 bp in length
* 37324 37423: gap of 100 bp
* 37424 42575: contig of 5152 bp in length
* 42576 42675: gap of 100 bp
* 42676 79708: contig of 37033 bp in length
* 79709 79808: gap of 100 bp
* 79809 86724: contig of 6916 bp in length
* 86725 86824: gap of 100 bp
* 86825 109536: contig of 22812 bp in length
* 109637 109736: gap of 100 bp
* 109737 139954: contig of 30218 bp in length
* 139955 140054: gap of 100 bp
* 140055 145722: contig of 5668 bp in length.
*
*
FEATURES
source
1.145722
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-21H2"
/clone_lib="RPCT-11 Human Male BAC"
1..10503
/note="assembly_fragment
clone_end:SP6
vector_side:left"
10504..10603
/estimated_length=100
10604..12771
/note="assembly_fragment"
12772..12871
/estimated_length=100
12872..14989
/note="assembly_fragment"
14990..15089
/estimated_length=100
15090..16911
/note="assembly_fragment"
16912..17011
/estimated_length=100
17012..19684
/note="assembly_fragment"

```

	gap	/estimated_length=100 19685..19784	
	misc_feature	/note="assembly_fragment" 19785..22088	
	gap	/estimated_length=100 22089..22188	
	misc_feature	/note="assembly_fragment" 22189..25445	
	gap	/estimated_length=100 25446..25545	
	misc_feature	/note="assembly_fragment" 25546..30628	
	gap	/estimated_length=100 30629..30728	
	misc_feature	/estimated_length=100 30729..37323	
	gap	/note="assembly_fragment" 37324..37423	
	misc_feature	/estimated_length=100 37424..42575	
	gap	/note="assembly_fragment" 42576..42675	
	misc_feature	/estimated_length=100 42676..79708	
	gap	/note="assembly_fragment" 79709..79808	
	misc_feature	/estimated_length=100 79809..86724	
	gap	/note="assembly_fragment" 86725..86824	
	misc_feature	/estimated_length=100 86825..109636	
	gap	/note="assembly_fragment" 109637..109736	
	misc_feature	/estimated_length=100 109737..139954	
	gap	/note="assembly_fragment" 139955..140054	
	misc_feature	/estimated_length=100 140055..145722	
		/note="assembly_fragment"	
		clone end:T7	
		vector_side:right"	
ORIGIN			
Query Match	87.2%; Score 21.8; DB 14;	Length 145722;	
Best Local Similarity	92.0%; Pred. No. 3.3e+02;		
Matches 23; Conservative	0; Mismatches 2;	Indels 0; Gaps 0;	
OY	1 AAAAAAAAACTAAAGCTTGATCCTT 25		
Ddb	46058 AGAAAAAAAATTAAACCTTGATCCTT 46034		
RESULT 15			
LOCUS AC073494			
DEFINITION Homo sapiens chromosome X clone RP11-1P14,		linear HTG 08-JAN-2003	
ACTION 11 unordered pieces.		WORKING DRAFT SEQUENCE,	
ACCESSION AC073494			
VERSION AC073494.9 GI:20335720			
KEYWORDS ATGS; HTGS_PHASEI; HTGS_DRAFI.			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
	Mammalia Eutheria Chordata Craniata Vertebrata Euteleostomi;		
	Eukaryota Metazoa Chordata Craniata Vertebrata Euteleostomi;		
	Hominidae Homo.		
REFERENCE 1 (bases 1 to 160936)			
AUTHORS Mizny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshrooke,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,J., Bimaga,K., Blankenburg,X., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Butrell,K.I., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,			

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carrillo, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudish, S., Karlsson, E., Keshi, A., Khan, U., King, L., Korvah, J., Kovar, C., Kravtsov, J., Kureishi, A., Landry, N., Deal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteberg, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mashiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S., Ogun, M., Okunodu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitani, N., Sison, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Syarik, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Teitrod, B., Thomas, N., Thomas, S., Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wiczyski, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Unpublished
2 (bases 1 to 160936)
Worley, K.C.
Direct Submission
Submitted (19-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 160936)
Worley, K.C.
Direct Submission
Submitted (08-JUN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:15789212.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: HBSP
Center clone name: RP11-1F14

Summary Statistics
Sequencing vector: M13
Chemistry: Dye-primer Bodipy: 44% of reads
Chemistry: Dye-terminator Big Dye: 56% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 152384 bases at least Q40
Consensus quality: 154433 bases at least Q30
Consensus quality: 155875 bases at least Q20
Estimated insert size: 158401; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2079: contig of 2079 bp in length
* 2080 2179: gap of unknown length
* 2180 4253: contig of 2074 bp in length
* 4254 4353: gap of unknown length
* 4354 7577: contig of 3224 bp in length
* 7578 7677: gap of unknown length
* 7678 13371: contig of 5694 bp in length
* 13372 13471: gap of unknown length
* 13472 25082: contig of 11611 bp in length
* 25083 25182: gap of unknown length
* 25183 34508: contig of 9326 bp in length
* 34509 34608: gap of unknown length
* 34609 51601: contig of 16993 bp in length
* 51602 51701: gap of unknown length
* 51702 66858: contig of 15157 bp in length
* 66859 66958: gap of unknown length
* 66959 97199: contig of 30241 bp in length
* 97200 97299: gap of unknown length
* 97300 127960: contig of 30661 bp in length
* 127961 128060: gap of unknown length
* 128061 160936: contig of 32876 bp in length.
Location/Qualifiers
1. 160936
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-1F14"
2080. 2179
/estimated_length=unknown
4254. 4353
/estimated_length=unknown
7578. 7677
/estimated_length=unknown
13372. 13471
/estimated_length=unknown
25083. 25182
/estimated_length=unknown
34509. 34608
/estimated_length=unknown
51602. 51701
/estimated_length=unknown
66859. 66958
/estimated_length=unknown
97200. 97299
/estimated_length=unknown
127961. 128060
/estimated_length=unknown

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 160936;
Best Local Similarity 92.0%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAAGCTTATCTT 25
Db 156878 AAAAAAAAAATTAAAGCTTATCTT 156902

RESULT 16
AC145988/c 162223 bp DNA linear PRI 18-DEC-2003
DEFINITION Pan troglodytes BAC clone RP43-166B1 from 7, complete sequence.
ACCESSION AC145988
VERSION AC145988.3 GI:38424259
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Hominidae; Pan.
1 (bases 1 to 162223)
Hodges, J. and Haglund, K.
The sequence of Pan troglodytes BAC clone RP43-166E1
Unpublished (2001)
2 (bases 1 to 162223)
Wilson, R.
Sequencing of Pan troglodytes
Unpublished (2001)
3 (bases 1 to 162223)
Wilson, R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 162223)
Wilson, R.K.
Direct Submission
Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 162223)
Wilson, R.K.
Direct Submission
Submitted (19-NOV-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 162223)
Wilson, R.
Direct Submission
Submitted (18-DEC-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 19, 2003 this sequence version replaced gi:38154103.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics -----
Center project name: C_P716B01

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RCI-13 BAC library has been constructed by Chung-Li Shu. DNA
was isolated from white blood cells obtained from a male chimpanzee
(Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The
clone and detailed information can be obtained from Resgen
(http://www.resgen.com) or Pieter de Jong and co-workers at
http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1..162223
/organism="Pan troglodytes"

```

/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"
/cclone="RP43-166E1"
/cclone_1lb="RPCI-43"
92396..92779
/note=Sequence derived from one plasmid subclone."
-----
ORIGIN
unauure
Query Match      87.2%; Score 21.0; DB 8; Length 162223;
Best Local Similarity 92.0%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy      1 AAAAAAAAACTAAGCTTGATCTT 25
          |||||
Db       149243 AAAAAAAAAAAAGCTTGATCTT 149219
-----
RESULT 17
AC015648/c
LOCUS
DEFINITION
Homo sapiens clone RP11-46C18, WORKING DRAFT SEQUENCE, 18 unordered
pieces.
AC015648 GI:7329258
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
Eukaryotic; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 166019)
Birren,B., Linton,L., Nusbaum,C. and Landier,E.
Homo sapiens, clone RP11-46C18
Unpublished
2 (bases 1 to 166019)
Birren,B., Linton,L., Nusbaum,C., Landier,E., Allen,N., Anderson,M.,
Baldwin,J., Barrn,A., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cook,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J.,
Galagan,J.C., Gardyna,S., Grant,G., Hages,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatez,A., Klein,J.,
Lehoczeky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,D.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,U.,
Tyfanie,S., Tittrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.A.
DIRECT Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6730920.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
-----
TITLE
JOURNAL
COMMENT
-----
Center project name: l1234
Center clone name: 46_C18
----- Summary Statistics
Sequencing vector: MJ3; M77915; 100% of reads
Chemistry: Dye-primer-amersham; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156140 bases at least Q40
Consensus quality: 160433 bases at least Q30

```

Consensus quality: 162210 bases at least Q20
 Insert size: 170000; agarose-fp
 Insert size: 164319; sum-of-contigs
 Quality coverage: 4.9 in Q20 bases; agarose-fp
 Quality coverage: 5.0 in Q20 bas.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1564: contig of 1564 bp in length
 1565 1664: gap of 100 bp
 * 1665 3109: contig of 1445 bp in length
 * 3110 3509: gap of 100 bp
 * 3210 6556: contig of 3347 bp in length
 * 6557 9048: gap of 100 bp
 * 9049 9148: contig of 2392 bp in length
 * 9149 10847: gap of 100 bp
 * 10848 10947: contig of 1699 bp in length
 * 10948 14273: gap of 100 bp
 * 14274 14374: contig of 3326 bp in length
 * 14374 18490: contig of 4117 bp in length
 * 18491 18590: gap of 100 bp
 * 18591 22982: contig of 4392 bp in length
 * 22983 23082: gap of 100 bp
 * 23083 27931: contig of 4749 bp in length
 * 27932 32167: contig of 4236 bp in length
 * 32168 32267: gap of 100 bp
 * 32268 40215: contig of 7948 bp in length
 * 40216 40315: gap of 100 bp
 * 40316 47861: contig of 7546 bp in length
 * 47862 47961: gap of 100 bp
 * 47962 62372: contig of 1441 bp in length
 * 62373 62472: gap of 100 bp
 * 62473 73564: contig of 11092 bp in length
 * 73565 73664: gap of 100 bp
 * 73665 88229: contig of 14565 bp in length
 * 88230 88329: gap of 100 bp
 * 88330 108403: contig of 20074 bp in length
 * 108404 108503: gap of 100 bp
 * 108504 134596: contig of 26093 bp in length
 * 134597 134697: gap of 100 bp
 * 134697 166019: contig of 31323 bp in length.

FEATURES

source 1..166019
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-46C18"
 /clone_id="RPC1-11 Human Male BAC"
 misc_feature 1..1564
 /note="assembly_fragment"
 gap 1565..1664
 /estimated_length=100
 misc_feature 1665..3109
 /note="assembly_fragment"
 gap 3110..3209
 /estimated_length=100
 misc_feature 3210..6556
 /note="assembly_fragment"
 gap 6557..6656
 /estimated_length=100
 misc_feature 6657..9048
 /note="assembly_fragment"
 gap 9049..9148
 /estimated_length=100
 misc_feature 9149..10847
 /note="assembly_fragment"

clone_end:8P6
 vector_side:left"
 10848..10947
 /estimated_length=100
 10948..14273
 /note="assembly_fragment"
 14274..14373
 /estimated_length=100
 14374..18490
 /note="assembly_fragment"
 18491..18590
 /estimated_length=100
 18591..22982
 /note="assembly_fragment"
 22983..23082
 /estimated_length=100
 23083..27831
 /note="assembly_fragment"
 27832..27931
 /estimated_length=100
 27932..32167
 /note="assembly_fragment"
 32168..32267
 /estimated_length=100
 32268..40215
 /note="assembly_fragment"
 40216..40315
 /estimated_length=100
 40316..47861
 /note="assembly_fragment"
 47862..47961
 /estimated_length=100
 47962..62372
 /note="assembly_fragment"
 62373..62472
 /estimated_length=100
 62473..73564
 /note="assembly_fragment"
 73565..73664
 /estimated_length=100
 73665..88229
 /note="assembly_fragment"
 88230..88329
 /estimated_length=100
 88330..108403
 /note="assembly_fragment"
 108404..108503
 /estimated_length=100
 108504..134596
 /note="assembly_fragment"
 clone_end:17
 vector_side:right"
 134597..134696
 /estimated_length=100
 134697..166019
 /note="assembly_fragment"

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 166019;
 Best Local Similarity 92.0%; Pred. No. 3.1e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
 Db 77015 AGAAAAAAAAATAAGCTTGATCTT 76991

RESULT 18
 CR936342/c 187909 bp DNA linear HTG 25-APR-2005
 LOCUS CR936342
 DEFINITION Danio rerio clone CH211-215D19, *** SEQUENCING IN PROGRESS ***
 ACCESSION CR936342

```

VERSION      CR936342.5  GI:62896454
KEYWORDS     HTG: HTGS PHASE1.
SOURCE       Danio rerio (zebrafish)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
              1 (bases 1 to 187909)
REFERENCE    McLay,K.
AUTHORS      Direct Submission
TITLE        Submitted (24-APR-2005) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              zfish-help@sanger.ac.uk
              zfish-help@sanger.ac.uk
              On Apr 25, 2005 this sequence version replaced gi:62866421.
COMMENT      ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: zfish-help@sanger.ac.uk
              ----- Project Information
              Center project name: zc215D19
              ----- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Chemistry: Dye-terminator; 100% of reads
              Consensus quality: 182145 bases at least Q40
              Consensus quality: 183196 bases at least Q30
              Consensus quality: 184102 bases at least Q20
              Insert size: 186609; sum-of-contigs
              Insert size: 128222; 45.1% exons; agarose-gel
              Quality coverage: 8.09x in Q20 bases; sum-of-contigs Quality
              coverage: 11.78x in Q20 bases; agarose-gel
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 14 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              *
              1      13614: contig of 13614 bp in length
              *      13714: gap of 100 bp
              *      13715      17307: contig of 3593 bp in length
              *      17308      17407: gap of 100 bp
              *      17408      22487: contig of 5080 bp in length
              *      22488      22587: gap of 100 bp
              *      22588      35614: contig of 13027 bp in length
              *      35615      35714: gap of 100 bp
              *      35715      51458: contig of 15744 bp in length
              *      51459      51558: gap of 100 bp
              *      51559      62770: contig of 11212 bp in length
              *      62771      62870: gap of 100 bp
              *      62871      72450: contig of 9580 bp in length
              *      72451      72550: gap of 100 bp
              *      72551      88008: contig of 15458 bp in length
              *      88009      88108: gap of 100 bp
              *      88109      90243: contig of 2135 bp in length
              *      90244      90343: gap of 100 bp
              *      90344      101483: contig of 11140 bp in length
              *      101484      101583: gap of 100 bp
              *      101584      104058: contig of 2475 bp in length
              *      104059      104158: gap of 100 bp
              *      104159      107404: contig of 3246 bp in length
              *      107405      107504: gap of 100 bp
              *      107505      127836: contig of 20332 bp in length
              *      127837      127936: gap of 100 bp
              *      127937      187909: contig of 59973 bp in length.
              *
              Location/Qualifiers
              1..187909
              /organism="Danio rerio"
              /mol_type="genomic DNA"
              /db_xref="taxon:7955"
              /clone="CH211-215D19"

```

	misc_feature	/clone_1fb="CHORI-211"
		1..13614
		/note="assembly_fragment:01008
		clone_end:T7
		vector_side:left"
	misc_feature	13715..17307
		/note="assembly_fragment:00023
		fragment_chain:1"
	misc_feature	17408..22487
		/note="assembly_fragment:00071
		fragment_chain:1"
	misc_feature	22588..35614
		/note="assembly_fragment:00236
		fragment_chain:1"
	misc_feature	35715..51458
		/note="assembly_fragment:00629
		fragment_chain:1"
	misc_feature	51559..62770
		/note="assembly_fragment:00478
		fragment_chain:1"
	misc_feature	62871..72450
		/note="assembly_fragment:00353
		fragment_chain:2"
	misc_feature	72551..88008
		/note="assembly_fragment:00805
		fragment_chain:2"
	misc_feature	88109..90243
		/note="assembly_fragment:00013"
	misc_feature	90344..101483
		/note="assembly_fragment:00122"
	misc_feature	101584..104058
		/note="assembly_fragment:00035
		fragment_chain:3"
	misc_feature	104159..107404
		/note="assembly_fragment:00047
		fragment_chain:3"
	misc_feature	107505..127836
		/note="assembly_fragment:01242
		fragment_chain:3"
	misc_feature	127937..187909
		/note="assembly_fragment:01477
		fragment_chain:3
		clone_end:SP6
		vector_side:right"
ORIGIN		
Query Match	87.2%	Score 21.8; DB 14;
Best Local Similarity	92.0%;	Pred. No. 3e+02; Length 187909;
Matches 23; Conservative	0; Mismatches 2;	Indels 0; Gaps 0;
OY	1 AAAAAAAAACTAAGCTTGATCTT 25 	
Db	178070 AAAAAAAACTTAAGTTTATCTT 178046	
RESULT 19		
AL603662/c		
LOCUS	201275 bp DNA linear ROD 16-APR-2005	
DEFINITION	Mouse DNA sequence from clone RP23-396M19 on chromosome 11 Contains the 3' end of the Myh10 gene for non-muscle myosin heavy chain 10, the NdelI gene for nuclear distribution gene B-like homolog 1 (A. nidiulans), three genes for novel proteins (9930039A1IRik, 1810012H1IRik and D130071N09), the Rpl26 gene for ribosomal protein L26, gene Oppo1-pending (oppo 1), gene Ranl guanine nucleotide release factor (Rangnrf-pending) and three Cps islands, complete sequence.	
ACCESSION	AL603662	
VERSION	AL603662.11 GI:17736577	
KEYWORDS	HNG; 1810012H1IRik; 9930039A1IRik; D130071N09; Myh10; NdelI; oppo 1; Oppo1-pending; Rangnrf-pending; Rpl26.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mus musculus	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201275)
Pearce, A.
Direct Submission
Submitted (04-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgehire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Dec 13, 2001 this sequence version replaced gi:17381379.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.; EMBL; Sw.; SWISSPROT; Tr.; TrEMBL; Wp.; WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep -----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
RP23-396M19 is from the RP23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers
1. 201275
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-396M19"
/clone_1fb="RP23-396M19-23"
join(Al645644.8:101766..102014,Al645644.8:109042..109417,
Al645644.8:131152..131308,Al645644.8:144220..144247,
Al645644.8:155149..155251,Al645644.8:161774..161866,
Al645644.8:169578..169641,Al645644.8:170485..170583,
Al645644.8:172343..172486,Al645644.8:172670..172765,
Al645644.8:174068..174186,Al645644.8:174873..175025,
Al645644.8:178031..178204,Al645644.8:180716..180889,
Al645644.8:181704..181818,Al645644.8:182938..193131,
Al645644.8:193215..193336,Al645644.8:192938..194828,
Al645644.8:195028..195188,Al645644.8:194759..194828,
Al645644.8:195028..195188,Al645644.8:194759..194828,
3848..4054,6615..6752,7127..7250,8373..8544,9566..9778,
13842..13986,19350..19562,17325..17429,18087..18239,
18618..18966,19350..19562,21137..21349,23635..23796,
25303..25431,26670..26758,28111..28234,28307..28515,
29479..29587,30892..31064,31437..33135)
/gene="Myh10"
/locus_tag="RP23-396M19.2-001"
join(Al645644.8:101766..102014,Al645644.8:109042..109417,
Al645644.8:131152..131308,Al645644.8:144220..144247,
Al645644.8:155149..155251,Al645644.8:161774..161866,
Al645644.8:169578..169641,Al645644.8:170485..170583,
Al645644.8:172343..172486,Al645644.8:172670..172765,
Al645644.8:174068..174186,Al645644.8:174873..175025,
Al645644.8:178031..178204,Al645644.8:180716..180889,
Al645644.8:181704..181818,Al645644.8:182938..193131,
Al645644.8:193215..193336,Al645644.8:192938..194828,
Al645644.8:195028..195188,Al645644.8:194759..194828,
Al645644.8:195028..195188,Al645644.8:194759..194828,
3848..4054,6615..6752,7127..7250,8373..8544,9566..9778,
13842..13986,19350..19562,17325..17429,18087..18239,
18618..18966,19350..19562,21137..21349,23635..23796,
25303..25431,26670..26758,28111..28234,28307..28515,
29479..29587,30892..31064,31437..33135)
/gene="Myh10"

gene

25303..25431,26670..26758,28111..28234,28307..28515,
29479..29587,30892..31064,31437..33135)
/gene="Myh10"
/locus_tag="RP23-396M19.2-001"
/product="myosin heavy chain 10, non-muscle"
/note="match: ESTs: BY005341.1
match: cDNAs: AB022023.1 AF133055.1 AK026977.1 AK029236.1
AK050066.1 AK045074.1 AK047340.1 AK052435.1 AK122578.1
M69181.1 U34304.1"
join(Al645644.8:109070..109417,Al645644.8:131152..131308,
Al645644.8:144220..144247,Al645644.8:155149..155251,
Al645644.8:155767..155796,Al645644.8:161774..161866,
Al645644.8:169578..169641,Al645644.8:170485..170583,
Al645644.8:172343..172486,Al645644.8:172670..172765,
Al645644.8:174068..174186,Al645644.8:174873..175025,
Al645644.8:178031..178204,Al645644.8:180716..180889,
Al645644.8:181704..181818,Al645644.8:185197..185259,
Al645644.8:192938..193131,Al645644.8:193215..193336,
Al645644.8:194759..194828,Al645644.8:195028..195188,
2128..2236,2335..2466,3848..4054,6615..6752,7127..7250,
8373..8544,9566..9778,13842..13986,1760..14966,
17325..17429,18087..18239,1618..1886,19350..19562,
21137..21349,23635..23796,25303..25431,26670..26758,
28111..28234,28307..28515,29479..29587,30892..31064,
31437..31583)
/gene="Myh10"
/locus_tag="RP23-396M19.2-002"
join(Al645644.8:109070..109417,Al645644.8:131152..131308,
Al645644.8:144220..144247,Al645644.8:155149..155251,
Al645644.8:155767..155796,Al645644.8:161774..161866,
Al645644.8:169578..169641,Al645644.8:170485..170583,
Al645644.8:172343..172486,Al645644.8:172670..172765,
Al645644.8:174068..174186,Al645644.8:174873..175025,
Al645644.8:178031..178204,Al645644.8:180716..180889,
Al645644.8:181704..181818,Al645644.8:185197..185259,
Al645644.8:192938..193131,Al645644.8:193215..193336,
Al645644.8:194759..194828,Al645644.8:195028..195188,
2128..2236,2335..2466,3848..4054,6615..6752,7127..7250,
8373..8544,9566..9778,13842..13986,1760..14966,
17325..17429,18087..18239,1618..1886,19350..19562,
21137..21349,23635..23796,25303..25431,26670..26758,
28111..28234,28307..28515,29479..29587,30892..31064,
31437..31583)
/locus_tag="RP23-396M19.2-002"
/product="myosin heavy chain 10, non-muscle"
/note="match: cDNAs: M93676.1 U15693.1 U73303.1"
join(Al645644.8:144220..144247,Al645644.8:155149..155251,
Al645644.8:155767..155796,Al645644.8:161774..161866,
Al645644.8:169578..169641,Al645644.8:170485..170583,
Al645644.8:172343..172486,Al645644.8:172670..172765,
Al645644.8:174068..174186,Al645644.8:174873..175025,
Al645644.8:178031..178204,Al645644.8:180716..180889,
Al645644.8:181704..181818,Al645644.8:185197..185259,
Al645644.8:192938..193131,Al645644.8:193215..193336,
Al645644.8:194759..194828,Al645644.8:195028..195188,
2128..2236,2335..2466,3848..4054,6615..6752,7127..7250,
8373..8544,9566..9778,13842..13986,1760..14966,
17325..17429,18087..18239,1618..1886,19350..19562,
21137..21349,23635..23796,25303..25431,26670..26758,
28111..28234,28307..28515,29479..29587,30892..31064,
31437..31581)
/gene="Myh10"
/locus_tag="RP23-396M19.2-002"
/product="myosin heavy chain 10, non-muscle"
/note="match: cDNAs: M93676.1 U15693.1 U73303.1"
join(Al645644.8:144220..144247,Al645644.8:155149..155251,
Al645644.8:155767..155796,Al645644.8:161774..161866,
Al645644.8:169578..169641,Al645644.8:170485..170583,
Al645644.8:172343..172486,Al645644.8:172670..172765,
Al645644.8:174068..174186,Al645644.8:174873..175025,
Al645644.8:178031..178204,Al645644.8:180716..180889,
Al645644.8:181704..181818,Al645644.8:185197..185259,
Al645644.8:192938..193131,Al645644.8:193215..193336,
Al645644.8:194759..194828,Al645644.8:195028..195188,
2128..2236,2335..2466,3848..4054,6615..6752,7127..7250,
8373..8544,9566..9778,13842..13986,1760..14966,
17325..17429,18087..18239,1618..1886,19350..19562,
21137..21349,23635..23796,25303..25431,26670..26758,
28111..28234,28307..28515,29479..29587,30892..31064,
31437..31581)
/codon_start=1
/protein_id="CA125527.1"
/db_xref="GI:56205559"
/db_xref="GOA:Q5SV64"
/db_xref="InterPro:IPR000048"
/db_xref="InterPro:IPR001609"

CDS

```

/db_xref="InterPro:IPR002017"
/db_xref="InterPro:IPR002928"
/db_xref="InterPro:IPR004009"
/db_xref="UniProt/TREMBL:Q5S64"
/translation="MAORTGLDEPERLYFVDRVAVYNPATQADWTAKLAVIIPSEHNG
FEASISKEERGDEVMVELAENGKAMKANKDOKNPPKFSVEMALTCINSEVHG
HNKORRYSGILYVSGFCVINYINPKLPIYENIEMVYKKKKKDNPHIYASIS
AYRCMLQDREPOSIICTSSESGAKTENTKTYOYLAHVASSHKKKKDNPHIQESKPY
KQGLEKQQLQANPILSPGAKTVKNDNSRFGEKTRINFDVGTGYVGNATYLYL
EKSRVQAKDERTHIFLYQLSGAGELKADLLLEGNNRFLPSNGIPIPGQDDKD
NFOETMAMHMGFSHEEILSMKVSSVLQKSTSPKKNRTDASVEMPTVYKIC
HLGNVMEFTRAILTPRIKVGQDVOKAQTEQADPFAVLAATYRLEFRMLVHRI
NKALDRTKQASFCIGLIDGAFETFEINSFEOLCINTNKKLQDLPHTMFTLIEOE
YORBCIENNFIDFGDLPDPCIDLIRPANNPQVALLBECFFPATTKTYEKKVOE
QGSHEKQPRQKQKADFCIITHAGKVDYKADENLKMOPNDNVAITLLHQSSDR
VAELMKDEIQTORASFYDSVSGHEPVDRVGLDQVTGWTETAFSGAYTKKMF
TVGOLYKESLTKMLATNTNPNFRCIIPNHEKAGKLDPHLVIDQRCNGVIGKIR
ICROGFPRIYVQEFORRYEILTPNAIPKGMWDQACERMIKALREMLYRIGQSK
IEFRAGVLAHLEERDLKITDIIIPFOACRGYLARKAPAKQOOLSKLYQRNCA
YIKLRHOMWRYFTYKPLLOYTROEERLQAKDESLVYKEQTKVGBELBEMKQO
QLEBKNTLQLOLQETELFAEAEERARLAKQLEELIHLDESREEEERQOIL
ONEKKQMAHIQDLEBQDDEEGARQKQLEKVTAEAKIKMEBEVILLBPONSKFIK
EKJLMEDRIACSSQLAESEBKAKLAKIRNKOEWIMISDLERLKEKERTOLEKAK
RLDGETTDLDQIAELQAYDELKVOLTKEEELQALAGDDDTLHKNNAKLYARE
LOAOIAELODPESEKSRKAKOKRDLSEELKLTLEBDPTLDITTAQOELPTKRE
OVBALKALDEDTNHEAQIDONKORATLAESELEBQDEQKPKANLEKQOLET
DNKELACVKNVQOVKASSEHRKLDQVOELHAKVSEGORLVSRLAEKANKQOEL
DNVSLTLEAEKKGIKFADKAGLSQLODTELQOETRQNLVSLRLEKANKQOEL
LOEOEERKAKNLEKOVLAQLOSLQATKKVDDDLGTISLEAKKKLLKDVALS
ORLEKVLAYDKLEKTKRLOOEELDLTVDDHORQIVSNLEKKEKTKDOLAEKGI
SARVAERDPAEAREKERTKSLARALERLEKKEPERONQOLRMDMDISKD
DVGKVELEKSKALBOQVEBMTQLELEDELOATEDAKLRVSNQAMQAKQERD
LQTRDEQNEKRRLLKQVRELAELEBERQALAVASKMEIDLQLESLQEAAN
KARBEVILQRLQAKMDYOURELEEARASDEIFQSKESKIKLSAEILQOE
LASSERARRHAEDEDELADEIANGSAGSKALDRLERLAEIOLBELEBEOGME
LINDREKTTLOVDTLNTLEAERSAQAQNDARQKONKELEKAKQOELGAYKSK
FKATISALEAKTIGLEBOLBOEKERAKRANLVTPEKLEKIPQVQDEERHDOYK
EOMEKANRMQKQLEBAEAEATRANASRRKLOREILDATANEGLSREVSTLKNR

```

Query Match 87.2%; Score 21.8; DB 9; Length 201275;
 Best Local Similarity 92.0%; Pred. No. 2.9e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 25
 Db 184101 AAAAAAAAAACAAACCTTGATCTT 184077

RESULT 20
 AC137947 214295 bp DNA linear ROD 16-JUL-2005
 LOCUS Mus musculus chromosome 5, clone RP23-36213, complete sequence.
 DEFINITION AC137947
 ACCESSION AC137947 GI:70909443
 VERSION HTG
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurionathi; Muridae; Murinae; Mus.
 1 (bases 1 to 214295)
 2 (bases 1 to 214295)
 2 (bases 1 to 214295)
 Birren,B., Nussbaum,C. and Lander,B.
 Unpublished
 Mammalian chromosome 5, clone RP23-36213
 Birren,B., Nussbaum,C., Lander,B., Ali,A., Allen,N., Anderson,S.,
 Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhgalter,B.,
 Camarata,D., Chang,D., Charazro,B., Choepel,Y., Collimore,A.,
 Cooke,P., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
 Gargay,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatae,A., Kelle,C., Landers,T., Levine,R.,

TITLE
 JOURNAL
 REFERENCES
 AUTHORS
 Submitted (05-DEC-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 214295)
 Birren,B., Nussbaum,C., Lander,B., Abouelleil,A., Allen,N.,
 Anderson,M., Anderson,S., Arachchi,H.M., Barina,N., Bastien,V.,
 Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,
 Choepel,Y., Collimore,A., Cook,A., Cooke,P., Cornu,B.,
 Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
 Erickson,J., Fato,S., Ferreira,P., FitzGerald,M., Gage,D.,
 Galagan,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
 Johnson,R., Jones,C., Kamat,A., Karatae,A., Kelle,C., Landers,T.,
 Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
 Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Menga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Notbu,C.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,C.,
 Phunkhang,P., Pierre,N., Rachupka,A., Ramaamy,U., Raymond,C.,
 Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R.,
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Stubbs,M., Talamae,J., Teste,S., Theodore,U.,
 Totham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
 Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemek,L.,
 Zimmer,A. and Zody,M.
 Direct Submission

TITLE
 JOURNAL
 REFERENCES
 AUTHORS
 Submitted (23-APR-2005) Broad Institute of MIT and Harvard, 320
 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 214295)
 Birren,B., Nussbaum,C., Lander,B., Abouelleil,A., Allen,N.,
 Anderson,M., Anderson,S., Arachchi,H.M., Barina,N., Bastien,V.,
 Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,
 Choepel,Y., Collimore,A., Cook,A., Cooke,P., Cornu,B.,
 Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
 Erickson,J., Fato,S., Ferreira,P., FitzGerald,M., Gage,D.,
 Galagan,D., Gargay,S., Graham,L., Grand-Pierre,N., Hafez,N.,
 Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
 Johnson,R., Jones,C., Kamat,A., Karatae,A., Kelle,C., Landers,T.,
 Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
 Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Menga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Notbu,C.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,C.,
 Phunkhang,P., Pierre,N., Rachupka,A., Ramaamy,U., Raymond,C.,
 Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupack,R.,
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Stubbs,M., Talamae,J., Teste,S., Theodore,U.,
 Totham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
 Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemek,L.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (16-JUL-2005) Broad Institute of MIT and Harvard, 320
 Charles Street, Cambridge, MA 02141, USA
 On Jul 16, 2005 this sequence version replaced gi:62868158.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Broad Institute of MIT and Harvard
 Center code: WBIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@broad.mit.edu
 ----- Project Information

Center project name: L26442
Center clone name: 362_I_3

FEATURES

SOURCE

```

1. 214295
   location=Qualifiers
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="5"
   /map="5"
   /clone="RP23-36213"
   /clone_id="RP23-23 Female Mouse BAC"
   complement(5..116)
   /rpt_family="ID_B1"
   584..635
   /rpt_family="(TG)n"
   complement(637..776)
   /rpt_family="B1_MM"
   862..1007
   /rpt_family="YREP_MM"
   complement(899..1022)
   /rpt_family="ID_B1"
   complement(1559..1706)
   /rpt_family="B4"
   1784..1958
   /rpt_family="B3A"
   complement(2053..2528)
   /rpt_family="MERVL_LTR"
   complement(2580..2668)
   /rpt_family="Lx9"
   complement(2802..2949)
   /rpt_family="UR1A"
   complement(2951..3057)
   /rpt_family="Lx9"
   3054..3093
   /rpt_family="Lx9"
   repeat_region
   3094..3268
   /rpt_family="B2_MM2"
   3269..3593
   /rpt_family="Lx9"
   complement(3627..3764)
   /rpt_family="B1_MM"
   3822..3942
   /rpt_family="B1_MM"
   3943..4013
   /rpt_family="B1_MM"
   complement(4059..4247)
   /rpt_family="B2_MM2"
   complement(4273..4344)
   /rpt_family="L1M4"
   complement(4367..4844)
   /rpt_family="OR1A-int"
   complement(4847..5163)
   /rpt_family="OR1A2"
   complement(5764..5883)
   /rpt_family="B1_MM"
   5930..6222
   /rpt_family="WTE"
   6371..6423
   /rpt_family="(GGA)n"
   6730..6916
   /rpt_family="B2_MM1"
   complement(7271..7392)
   /rpt_family="Tiger7"
   7931..8060
   /rpt_family="MIR"
   complement(8462..8624)
   /rpt_family="MSTC"
   9282..9372
   /rpt_family="PB1D10"
   9385..9468
   /rpt_family="(CGTG)n"
   9543..9608

```

```

repeat_region      /rpt_family="B4A"
                    9781..9989
repeat_region      /rpt_family="B3A"
                    10011..10155
repeat_region      /rpt_family="RSINE1"
                    complement(10379..10438)
repeat_region      /rpt_family="L2"
                    10469..10826
repeat_region      /rpt_family="(TG)n"
                    complement(10874..11189)
repeat_region      /rpt_family="B4A"
                    complement(11726..11793)
repeat_region      /rpt_family="RSINE1"
                    complement(12567..12764)
repeat_region      /rpt_family="B2_MM2"
                    12799..12941
repeat_region      /rpt_family="B1_MM"
                    12947..12972
repeat_region      /rpt_family="(CAA)n"
                    13256..13675
repeat_region      /rpt_family="WTC"
                    13691..13836
repeat_region      /rpt_family="RSINE1"
                    14094..14125
repeat_region      /rpt_family="AT_rich"
                    14497..14957
repeat_region      /rpt_family="MLT1D"
                    15840..16040
repeat_region      /rpt_family="(TG)n"
                    16054..16228
repeat_region      /rpt_family="(TG)n"

```

Query Match 87.2% Score 21.8; DB 9; Length 214295;
Best Local Similarity 92.0%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTGTATCTT 25
 Db 124275 AAAAAAAAAAGCAAAAGCTTATCTT 124299

RESULT 21
 AC156538/c
 LOCUS
 DEFINITION Bos taurus clone CH240-40C7, *** SEQUENCING IN PROGRESS ***, 16
 unsorted pieces.
 AC156538
 ACCESSION
 VERSION AC156538.2 GI:68265381
 KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 217253)
 AUTHORS Mzyny,D.Marie, Metzker,M.Lee, Abramzon,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Aisbrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biewlo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B.,
 Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dedertich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Davies,K.,
 Egan,A., Escotro,M., Eugene,C., Evans,C.A., Falle,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Frazer,C.M., Gabriel,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregiorgis,E., Geier,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idiebid, D., Jackson, A., Jakupson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpanty, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovits, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzenhwa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Maimoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawlawi, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pawlertak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L., P., Puzo, M., Quito, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Slason, I., Sitter, C. D., Smajic, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umant, K., Valas, R., Vera, V., Villaseana, D., Walron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weidenhausen, A., Weise, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Unpublished
2 (bases 1 to 217253)
Worley, K. C.

Direct Submission
Submitted (30-JAN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217253)
Cow Genome Sequencing Consortium.

Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 28, 2005 this sequence version replaced gi:58323918.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are oriented and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: FCNZ
Center clone name: CH240-40C7
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 211516 bases at least Q40
Consensus quality: 213263 bases at least Q30
Consensus quality: 214829 bases at least Q20
Estimated insert size: 218022; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES	source
1	2093: contig of 2093 bp in length
2094	2274: gap of 181 bp
2275	5872: contig of 3598 bp in length
5873	5922: gap of 50 bp
5923	8612: contig of 2690 bp in length
8613	8662: gap of 50 bp
8663	10874: contig of 2212 bp in length
10875	10974: gap of unknown length
10975	14902: contig of 3928 bp in length
14903	14952: gap of 50 bp
14953	73709: contig of 58757 bp in length
73710	73759: gap of 50 bp
73760	85214: contig of 11455 bp in length
85215	85264: gap of 50 bp
85265	121723: contig of 36459 bp in length
121724	121773: gap of 50 bp
121774	137416: contig of 15643 bp in length
137417	137466: gap of 50 bp
137467	139399: contig of 1933 bp in length
139400	139449: gap of 50 bp
139450	154541: contig of 15092 bp in length
154542	154591: gap of 50 bp
154592	161139: contig of 6548 bp in length
161140	161189: gap of 50 bp
161190	202540: contig of 41351 bp in length
202541	202590: gap of 50 bp
202591	204300: contig of 1710 bp in length
204301	204350: gap of 50 bp
204351	215980: contig of 11630 bp in length
215981	216080: gap of unknown length
216081	217253: contig of 1173 bp in length.
Location/Qualifiers	
1..217253	/organism="Bos taurus"
	/mol_type="genomic DNA"
	/db_xref="taxon:9913"
	/clone="CH240-40C7"
2094..2274	/estimated_length=181
5873..5922	/estimated_length=50
8613..8662	/estimated_length=50
10875..10974	/estimated_length=unknown
14903..14952	/estimated_length=50
73710..73759	/estimated_length=50
85215..85264	/estimated_length=50
121724..121773	/estimated_length=50
137417..137466	/estimated_length=50
139400..139449	/estimated_length=50
154542..154591	/estimated_length=50
161140..161189	/estimated_length=50
202541..202590	/estimated_length=50

```

gap                               /estimated_length=50
gap                               204301..204350
gap                               /estimated_length=50
gap                               215981..216080
gap                               /estimated_length=unknown

ORIGIN

Query Match                      87.2%; Score 21.8; DB 14; Length 217253;
Best Local Similarity 92.0%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTGATCTT 25
    |||||
Db 206242 AATAAAAAATTAAGCTTGAATCTT 206218

RESULT 22
CR931978 218875 bp DNA linear HTG 01-JUN-2005
LOCUS Danio rerio clone CH211-21306, WORKING DRAFT SEQUENCE, 7 unordered
DEFINITION pieces.
ACCESSION CR931978.5 GI:66863539
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Danio rerio (zebrafish)
SOURCE Danio rerio
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Osteichthyes;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 218875)
McLaren, S.
Direct Submission
Submitted (30-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
fish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 1, 2005 this sequence version replaced gi:5851924.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: fish-help@sanger.ac.uk
----- Project Information
Center project name: zc21306
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Consensus quality: 216050 bases at least Q40
Consensus quality: 216651 bases at least Q30
Consensus quality: 217204 bases at least Q20
Insert size: 218275; sum-of-contigs
Insert size: 205436; 4.7% error; agarose-fp
Quality coverage: 11.34x in Q20 bases; sum-of-contigs Quality
coverage: 12.57x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 26809: contig of 26809 bp in length
* 26910: gap of 100 bp
* 26910 68268: contig of 41359 bp in length
* 68269: gap of 100 bp
* 68369 167000: contig of 98632 bp in length
* 167001: gap of 100 bp
* 167101 171350: contig of 4250 bp in length
* 171351 171450: gap of 100 bp
* 171451 179079: contig of 7629 bp in length
* 179080: gap of 100 bp
* 179180 181788: contig of 2609 bp in length

```

```

FEATURES
    source
        * 181789 181888: gap of 100 bp
        * 181889 218875: contig of 36987 bp in length.
        Location/Qualifiers
            1..218875
                /organism="Danio rerio"
                /mol_type="genomic DNA"
                /db_xref="taxon:7955"
                /clone="CH211-21306"
                /clone_1id="CHOR1-211"
                1..26809
                    /note="assembly fragment:00079
                    fragment chain:1
                    clone end:T7
                    vector_side:left"
                26910..68268
                    /note="assembly fragment:01323
                    fragment chain:1"
                68369..167000
                    /note="assembly fragment:02099
                    fragment chain:1"
                167101..171350
                    /note="assembly fragment:00022"
                171451..179079
                    /note="assembly fragment:00053"
                179180..181788
                    /note="assembly fragment:00036
                    fragment chain:2
                    181889..218875
                    /note="assembly fragment:00624
                    fragment chain:2
                    clone end:SP6
                    vector_side:right"

ORIGIN

Query Match                      87.2%; Score 21.8; DB 14; Length 218875;
Best Local Similarity 92.0%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTGATCTT 25
    |||||
Db 175113 AAAAAAAAACTAAAGCTTATCTT 175089

RESULT 23
AC126071 227741 bp DNA linear ROD 31-AUG-2003
LOCUS Rattus norvegicus 10 BAC CH230-209B21 (Children's Hospital Oakland
DEFINITION Research Institute) complete sequence.
ACCESSION AC126071
VERSION AC126071.5 GI:3436593
KEYWORDS HTG.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murine; Rattus.
1 (bases 1 to 227741)
Muzny, D., Warren, W., Metzker, M., Lee, A., Adams, C., Alder, J.,
Allen, C., Allen, H., Alibekov, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blych, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, B.,
Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
Devila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Diaper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Davies, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

```

	Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hui, K., Hume, J., Idlibid, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Louiseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawlinay, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Minner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okunolu, G., Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poidexter, A., Popovic, D., Primus, E., Pul, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 227741)		
AUTHORS	Worley, K. C.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	3 (bases 1 to 227741)		
AUTHORS	Rat Genome Sequencing Consortium.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
REFERENCE	4 (bases 1 to 227741)		
AUTHORS	Worley, K. C.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-AUG-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
COMMENT	On Aug 31, 2003 this sequence version replaced gi:25138858. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.		
FEATURES	Location/Qualifiers		
SOURCE	1. 227741		
	/organism="Rattus norvegicus"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:10116"		
	/chromosome="10"		
	/clone="CH230-209821"		
	complement(397..491)		
	/rpt_family="PB1D9"		
	complement(508..683)		
	/rpt_family="B3"		
	712..854		
	/rpt_family="B1_Mur3"		
	1118..1165		
	/rpt_family="ATMG.n"		
	repeat_region	complement(1166..1254)	
		/rpt_family="ID_Rn"	
	repeat_region	complement(1273..1431)	
		/rpt_family="RSINE1"	
	repeat_region	complement(1999..2088)	
		/rpt_family="ID_Rn"	
	repeat_region	3586..3656	
		/rpt_family="(CA)n"	
	repeat_region	6415..6510	
		/rpt_family="ID_Rn"	
	repeat_region	6950..7065	
		/rpt_family="PB1D7"	
	repeat_region	7404..7437	
		/rpt_family="(TG)n"	
	repeat_region	7454..7545	
		/rpt_family="PB1D10"	
	repeat_region	7555..7650	
		/rpt_family="ID_Rn"	
	repeat_region	7935..7994	
		/rpt_family="(A)n"	
	repeat_region	7970..8117	
		/rpt_family="RSINE1"	
	repeat_region	complement(10486..10670)	
		/rpt_family="B2_Rn"	
	repeat_region	21355..21399	
		/rpt_family="GC_rich"	
	repeat_region	21520..21644	
		/rpt_family="(CCG)n"	
	repeat_region	22636..22819	
		/rpt_family="B2_Rn"	
	repeat_region	22835..22891	
		/rpt_family="B4"	
	repeat_region	22857..22894	
		/rpt_family="B1_Mur3"	
	repeat_region	23022..23207	
		/rpt_family="B2_Rn"	
	repeat_region	23256..23346	
		/rpt_family="PB1D7"	
	repeat_region	23406..23502	
		/rpt_family="PB1D9"	
	repeat_region	23623..23654	
		/rpt_family="(A)n"	
	repeat_region	complement(23929..24117)	
		/rpt_family="B2_Rn"	
	repeat_region	24167..24305	
		/rpt_family="B3A"	
	repeat_region	25092..25295	
		/rpt_family="B3"	
	repeat_region	25322..25459	
		/rpt_family="B1_Mur1"	
	repeat_region	25460..25498	
		/rpt_family="GA-rich"	
	repeat_region	25707..25798	
		/rpt_family="B3A"	
	repeat_region	26277..26386	
		/rpt_family="B3"	
	repeat_region	26412..26594	
		/rpt_family="B2_Rn"	
	repeat_region	26863..26885	
		/rpt_family="AT_rich"	
	repeat_region	complement(26933..27079)	
		/rpt_family="B1_Mur4"	
	repeat_region	28345..28693	
		/rpt_family="RLTR30"	
	repeat_region	28750..28821	
		/rpt_family="B1F"	
	repeat_region	28822..28893	
		/rpt_family="ID_Rn"	
	repeat_region	28894..28914	
		/rpt_family="(TG)n"	
	repeat_region	29197..29213	
		/rpt_family="AT_rich"	
	repeat_region	29214..29405	

```

repeat_region      /rpt family="B2_Rn"
                    29406..29419
repeat_region      /rpt family="AT_rich"
                    29420..29524
repeat_region      /rpt family="HCl_Rn"
                    29564..29737
repeat_region      /rpt family="B3A"
                    29797..30099
repeat_region      /rpt family="(CA)n"
                    complement(30112..30275)
repeat_region      /rpt family="B2_Rn"
                    30318..30378
repeat_region      /rpt family="(GCA)n"
                    30379..30748
repeat_region      /rpt family="RMR4B"
                    30772..30867
repeat_region      /rpt family="ID_Rn"
                    /rpt family="RMR3B"
                    complement(30868..30889)
repeat_region      /rpt family="RMR3B"
                    30890..30981
repeat_region      /rpt family="ID_Rn"
                    complement(30982..31177)
repeat_region      /rpt family="RMR3B"
                    31178..31281
repeat_region      /rpt family="ID_Rn"
                    complement(31282..31556)

```

```

Query Match      87.2% Score 21.8; DB 9; Length 227741;
Base Local Similarity 92.0%; Pred.No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Cy 1 AAAAAAAAAAAGTTGATCTT 25
Db 23640 AAAAAAAAAAAGTTGATCTT 23664

```

```

RESULT 24
AC158901      230880 bp DNA linear HTG 22-APR-2005
LOCUS      Mus musculus chromosome 15 clone RP23-77K13 map 15, *** SEQUENCING
DEFINITION      IN PROGRESSES ***, 11 unordered pieces.
ACCESSION      AC158901
VERSION      AC158901.4 GI:62860411
KEYWORDS      HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 230880)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE      Mus musculus chromosome 15, clone RP23-77K13
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 230880)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J.,
Choepeil,Y., Collamore,A., Cook,A., Cooke,P., Corum,B.,
Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galgan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
Maclean,C., Macdonald,P., Menes,L., Mihova,T., Mlenga,V.,
McCarthy,M., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J.,
Tophan,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.

```

```

TITLE      Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
DIRECT SUBMISSION      Submitted (24-MAR-2005) Broad Institute of MIT and Harvard, 320
JOURNAL      Charles Street, Cambridge, MA 02141, USA
REFERENCE      3 (bases 1 to 230880)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J.,
Choepeil,Y., Collamore,A., Cook,A., Cooke,P., Corum,B.,
Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galgan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
Maclean,C., Macdonald,P., Menes,L., Mihova,T., Mlenga,V.,
McCarthy,M., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J.,
Tophan,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.

```

```

TITLE      Direct Submission
JOURNAL      Submitted (22-APR-2005) Broad Institute of MIT and Harvard, 320
COMMENT      Charles Street, Cambridge, MA 02141, USA
On Apr 22, 2005 this sequence version replaced gi:62360737.
All repeats were identified using RepeatMasker:
Smt, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

Center: Broad Institute of MIT and Harvard
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L31328
Center clone name: 77_K_13
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 63601: contig of 63601 bp in length
* 63602 63701: gap of unknown length
* 63702 72918: contig of 9217 bp in length
* 72919 73018: gap of unknown length
* 73019 97625: contig of 24607 bp in length
* 97626 97725: gap of unknown length
* 97726 101462: contig of 3737 bp in length
* 101463 101562: gap of unknown length
* 101563 128930: contig of 27368 bp in length
* 128931 129030: gap of unknown length
* 129031 159922: contig of 40892 bp in length
* 159923 170022: gap of unknown length
* 170023 188971: contig of 18949 bp in length
* 188972 189071: gap of unknown length
* 189072 213374: contig of 24303 bp in length
* 213375 213474: gap of unknown length
* 213475 216601: contig of 3127 bp in length
* 216602 216701: gap of unknown length
* 216702 220589: contig of 3888 bp in length
* 220590 230880: gap of unknown length
* 230880: contig of 10191 bp in length.
Location/Qualifiers

```

FEATURES

```

source
1. 230860
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-77K13"
/clone_11b="RP23-77K13 Female Mouse BAC"
63602. .63701
/estimated_length=unknown
72919. .73018
/estimated_length=unknown
97626. .97725
/estimated_length=unknown
101463. .101562
/estimated_length=unknown
128931. .129030
/estimated_length=unknown
169923. .170022
/estimated_length=unknown
188972. .189071
/estimated_length=unknown
213375. .213474
/estimated_length=unknown
216602. .216701
/estimated_length=unknown
220590. .220689
/estimated_length=unknown

ORIGIN
Query Match 87.2%; Score 21.8; DB 14; Length 230860;
Best Local Similarity 92.0%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 61632 AAAAAAAAACTAAGCTTGATCTT 61656

RESULT 25
AC161390/c
LOCUS AC161390 236253 bp DNA linear HTG 01-JUL-2005
DEFINITION Bos taurus clone CH240-911L2, *** SEQUENCING IN PROGRESS ***, 30
uncloned pieces.
ACCESSION AC161390
VERSION AC161390.2 GI:68302897
KEYWORDS HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 236253)
Muzny,D,Marie, Metzker,M,lee, Abramzon,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsdorfs,S, Amin,A, Anguiano,D,
Anyalelechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barnstead,M, Barnstead,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Frazer,C,M, Gabisi,A, Ganter,R, Garcia,A, Garner,T, Garza,M,
Gebregorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M,
Hollins,B, Howells,S, Hulyk,S, Hume,T, Idlebird,D, Jackson,A,

```

```

TITLE JOURNAL
AUTHORS Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
AUTHORS Kowals,C, Kraft,C,L, Lebow,H, Levan,L, Lewis,L, Li,Z, Liu,J,
AUTHORS Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
AUTHORS Lorensheva,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J,
AUTHORS Maheshwari,M, Mahindartne,M, Mahmoud,M, Malloy,K, Mangum,A,
AUTHORS Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
AUTHORS Manthey,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
AUTHORS Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
AUTHORS Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L,
AUTHORS Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
AUTHORS Nwokediemo,O, Okonou,G, Olarunmagbon,A, Pal,S, Parke,K,
AUTHORS Paeternak,S, Paul,H, Perez,A, Perez,L, Piamkooch,C,
AUTHORS Plopper,F, Polindexter,A, Popovic,D, Primus,B, Pu,L,
AUTHORS Puazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
AUTHORS Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
AUTHORS Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
AUTHORS Sanders,W, Savery,G, Scherer,S, Scott,G, Shatman,S, Shen,H,
AUTHORS Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajd,D,
AUTHORS Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J,
AUTHORS Steimle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C,
AUTHORS Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umanu,K,
AUTHORS Valas,R, Vera,V, Villalana,D, Waldron,L, Walker,B, Wang,J,
AUTHORS Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
AUTHORS Williams,G, Willson,R, Wleczyk,R, Wooden,H, Worley,K,
AUTHORS Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V,
AUTHORS Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von
AUTHORS Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O,
AUTHORS Weinstock,G, and Gibbs,R,A.
Direct Submission
Unpublished
2 (bases 1 to 236253)
Worley,K,C.
Direct Submission
Submitted (13-MAY-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 236253)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2005 this sequence version replaced gi:63986587.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are oriented and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: FFWB
Center clone name: CH240-911L2
----- Summary Statistics
Assembly program: Atlas 3.0,
Consensus quality: 224955 bases at least Q40
Consensus quality: 227750 bases at least Q30
Consensus quality: 230049 bases at least Q20
Estimated insert size: 230265; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length

```

(see http://www.hgsc.bcm.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 30 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 4981: contig of 4981 bp in length
4982 5031: gap of 50 bp
5032 7523: contig of 2492 bp in length
7524 7573: gap of 50 bp
7574 35183: contig of 27610 bp in length
35184 35233: gap of 50 bp
35234 50751: contig of 1518 bp in length
50752 50801: gap of 50 bp
50802 65942: contig of 15141 bp in length
65943 65992: gap of 50 bp
65993 74714: contig of 8722 bp in length
74715 74764: gap of 50 bp
74765 74768: gap of 2744 bp in length
74769 77508: contig of 2744 bp in length
77509 77558: gap of 50 bp
77559 82120: contig of 4562 bp in length
82121 82220: gap of unknown length
82221 118027: contig of 35807 bp in length
118028 118077: gap of 50 bp
118078 130768: contig of 12691 bp in length
130769 130818: gap of 50 bp
130819 153112: contig of 22294 bp in length
153113 153162: gap of 50 bp
153163 161471: contig of 8309 bp in length
161472 161521: gap of 50 bp
161522 174530: contig of 13009 bp in length
174531 174580: gap of 50 bp
174581 178350: contig of 3770 bp in length
178351 178400: gap of 50 bp
178401 183715: contig of 5315 bp in length
183716 183765: gap of 50 bp
183766 185217: contig of 1552 bp in length
185218 185267: gap of 50 bp
185268 189137: contig of 3670 bp in length
189138 189187: gap of 50 bp
189188 190997: contig of 1810 bp in length
190998 191047: gap of 50 bp
191048 206593: contig of 15546 bp in length
206594 206693: gap of unknown length
206694 211029: contig of 4336 bp in length
211030 211079: gap of 50 bp
211080 212508: contig of 1429 bp in length
212509 212609: gap of unknown length
212610 217032: contig of 4424 bp in length
217033 217082: gap of 50 bp
217083 220073: contig of 2991 bp in length
220074 220398: gap of 325 bp
220399 221857: contig of 1459 bp in length
221858 222179: gap of 322 bp
222180 225156: contig of 2977 bp in length
225157 225256: gap of unknown length
225257 226324: contig of 1068 bp in length
226325 226424: gap of unknown length
226425 227750: contig of 1326 bp in length
227751 227850: gap of unknown length
227851 230239: contig of 2389 bp in length
230240 230339: gap of unknown length
230340 233635: contig of 3296 bp in length
233636 233735: gap of unknown length
233736 236253: contig of 2518 bp in length.
  
```

FEATURES
 SOURCE
 1. 236253
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"

```

gap      /clone="CH240-91112"
          4982..5031
          /estimated_length=50
gap      7524..7573
          /estimated_length=50
gap      35184..35233
          /estimated_length=50
gap      50752..50801
          /estimated_length=50
gap      50752..50801
          /estimated_length=50
gap      65943..65992
          /estimated_length=50
gap      74715..74764
          /estimated_length=50
gap      77509..77558
          /estimated_length=50
gap      82121..82220
          /estimated_length=unknown
gap      118028..118077
          /estimated_length=50
gap      130769..130818
          /estimated_length=50
gap      153113..153162
  
```

Query Match 87.2%; Score 21.8; DB 14; Length 236253;
 Best Local Similarity 92.0%; Pred. No. 2,8e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Oy 1 AAAAAAAAACTAAAGCTGATCTT 25
Db 155328 AAAAAAAAAATAAGCTGATCTT 155304
  
```

RESULT 26

AC106703

LOCUS

DEFINITION

AC106703.4

VERSION

KEYWORDS

SOURCE

ORGANISM

AUTHORS

AC106703 262198 bp DNA linear HTG 20-NOV-2002
 Rattus norvegicus clone CH230-213C17, *** SEQUENCING IN PROGRESS
 *, 5 unordered pieces.
 AC106703.4 GI:25139422
 HTG; HTG_PHASE1; HTG_DRAFT; HTG_ENRICHED.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
 1 (bases 1 to 262198)
 Muzny,D,Marle,Metzker,M,Lea,Abramson,S,Adams,C,Alder,J,
 Allen,C,Allen,H,Ashbrook,S,Amin,A,Angiano,D,
 Anyalebechi,V,Aoyagi,A,Ayodeji,M,Baca,E,Baden,H,
 Baldwin,D,Bandaranaike,D,Barber,M,Barnstead,M,Benahmed,F,
 Biewald,K,Blair,J,Blankenburg,K,Blyth,P,Brown,M,
 Bryant,N,Buhay,C,Burch,P,Burrell,K,Calderson,E,
 Cardenas,V,Carter,K,Cavazos,I,Ceasar,H,Center,A,
 Chacko,J,Chavez,D,Chen,G,Chen,R,Chen,Y,Chen,Z,
 Cleveland,C,Cockrell,R,Cox,C,Coyle,M,Cree,A,D'Souza,L,
 Devila,M,L,Davis,C,Davy-Carroll,L,De Anda,C,Dederich,D,
 Delgado,O,Denson,S,Deramo,C,Ding,Y,Dinh,H,Diyva,K,
 Draper,H,Dugan-Rocha,S,Dunn,A,Durbin,K,Duval,B,Eaves,K,
 Egan,A,Besotto,M,Eugene,C,Evans,C,A,Falls,T,Fan,G,
 Fernandez,S,Finley,M,Flagg,N,Forbes,L,Forster,M,Foster,P,
 Fraser,C,M,Gabriel,A,Ganta,R,Garcia,A,Garner,T,Gatza,W,
 Gbaggeorgis,E,Geer,K,Gill,R,Gredy,M,Guerra,W,Guevara,W,
 Guinatene,P,Haland,W,Hamil,C,Hamilton,C,Hamilton,K,
 Harvey,Y,Havlik,P,Hawes,S,Hladun,S,L,Hodgson,A,Hogues,M,
 Hernandez,R,Hines,S,Hladun,S,L,Hodgson,A,Hogues,M,
 Hollins,B,Howells,S,Huylk,S,Hume,J,Idelebird,D,Jackson,A,
 Jackson,L,Jacob,L,Jiang,H,Johnson,B,Johnson,R,Jolivet,A,
 Karpachy,S,Kelly,S,Kelly,S,Khan,Z,King,L,Koyar,C,
 Kowis,C,Kraft,C,D,Lebow,H,Leyan,J,Lewis,L,Li,Z,Liu,J,
 Liu,J,Liu,W,Liu,Y,London,P,Longacre,S,Lopez,J,
 Lorensuwa,L,Louisege,H,Lozano,R,J,Lu,X,Mac,J,
 Maheshwari,M,Mahindartine,M,Mahmoud,M,Mallory,K,Mangum,A,
 Mangum,B,Mapua,P,Martin,K,Martin,R,Martinez,E,

COMMENT

Center; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: meeki@gsc.riken.jp, URL: <http://range.gsc.riken.jp/>, Tel: 81-45-503-9625, Fax: 81-45-503-9586)
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). This clone is in a modified pBluescript vector. Please visit our web site (<http://range.gsc.riken.jp/>) for further details.

FEATURES

source

Location/Qualifiers
1..1681
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAFL16-77-D23"
/ecotype="Columbia"
/note="common name: thale cress"
1..1681
/gene="At3g21820"
61..1482
/gene="At3g21820"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD95436.1"
/db_xref="GI:62320769"
/translation="MDSYTKTDENFADVAALLAPLPLOQEFYFKLITRRNGIE
VKNNGITGKGVANSSEFDEDLIKDELIVIOHSSNKVDCLVCSFCFRFGSLTKQI
GRKLYFKNLGVSGCCDDSDSEEDCVKXNGEBQCGSSSHNTLPBGVVSVMNGEM
ALPHTDKFPLSPSCPGCCOAFYCSGCAAMWESSHSLCTGSESSISRALDE
FIKHANDNDFILAAKAIATILRYRLKAHYDKAKOSEPKOSLLEAMKPVISG
YKRWMPDIALPDVDPPTDEGAFPMOIKNLACTSLLEKIAIPKCEBALFSEIYCN
IIGMPELNNLDLVVAPSPVEDFYIYIDDLPAKSEETETITPFLDALGDEYSDCQQT
AFPPGCMNSCCPNAKAFREEDRGQAVIALRISKNEVTVIYIDELPYKER
QALLADYGFSCSKCKLEDSSSI"

gene
CDS

ORIGIN

Query Match 85.6%; Score 21.4; DB 15; Length 1681;
Best Local Similarity 95.7%; Pred. No. 2.2e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAAAAAAAACTAAAGCTTGATCTT 25
|||||
Db 1646 AAAAAAGAACTAAAGCTTGATCTT 1624

RESULT 28

AY059899

LOCUS

DEFINITION

complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AY059899 1639 bp mRNA linear PLN 04-NOV-2001
Arabidopsis thaliana Unknown protein (At3g21810; MSD21.12) mRNA,
complete cds.
AY059899
AY059899.1 GI:16649058
FLI CDNA.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 1639)
Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,U., Kamiya,A.,
Kawai,J., Kim,C., Lin,J., Liu,S.X., Natsumata,M., Pham,P.K.,
Sakano,H., Sakurai,T., Satou,M., Seki,M., Shim,P., Yamada,K.,
Shinozaki,K., Becker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (23-OCT-2001) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL CDNA: 'RIKEN Arabidopsis full-length cDNA'): Seki,M., Natsumata,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shim,P., Yamada,K., Becker,J., Theologis,A. and Davis,R.W.

Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers
1..1839
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="11"
/clone="RAFL08-19-N06"
/note="This clone is in pBluescript"
1..1839
/gene="At3g21810; MSD21.12"
272..1584
/gene="At3g21810; MSD21.12"
/codon_start=1
/product="unknown protein"
/protein_id="AAL24381.1"
/db_xref="GI:16649059"
/translation="MSSSSRYTKYKLCLEFNKTDCCSPNCTPAHNAELRRPGSSST
GRKHNMDSDLRNNHNDSDLRDLGQFSPERPSPIDRGRVQRPFGSDNMPFN
RKRKDYRENRFPDRDYAGLRYGNLIEPARDRGRKFGYNNVLEBQKDYMDYK
LMTDDKRLLEASVYERKAHEVDILTSRQELTOLDSEDCRRITSSKKFXYENF
LRAODDLKRESEARLQGNOLSTYLAGSEBNNDVGLDIVSDETNRNRTACDPHN
ELONTSSLSRKKHYVOYTKRPEVNSLIGRGEERKPKRPPCMNMLSSKYSSE
ESSANDEDTINSSSKEDWKKRRRRTSGTSATDKYLTNSMAAREDDVAABSEENP
EAAAGSPLSLPPEPRDAHVORDEDDVGVMEOKKAYDDDSV"

gene
CDS

ORIGIN

Query Match 85.6%; Score 21.4; DB 15; Length 1839;
Best Local Similarity 95.7%; Pred. No. 2.1e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAAAAAAAACTAAAGCTTGATCTT 25
|||||
Db 1689 AAAAAAGAACTAAAGCTTGATCTT 1711

RESULT 29

AB025634

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AB025634 59793 bp DNA linear PLN 14-FEB-2004
Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MSD21.
AB025634 B0000014
AB025634.1 GI:4589440
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1
Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
P1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
10819329
2 (bases 1 to 59793)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Direct Submission

CDS

CDS

NINPPTSTNFTTLQRLFRNOCILNKDLVLSGAHTIGVSHCSNMNRLYNSTTVKOD
 PSLDSOYVANKKANKCKSLINDNSTLEMDPGSSRSFSLSYRLVLRKRGGLQSDSALT
 TNSATLKYNLDLVNGSEKKFKPKAKSMKMRKRVKKGAGVIRTRCSVAGS"
 complement (31933..33372)
 /note="gene_id:MSD21.9"
 /codon_start=1
 /evidence=not_experimental
 /product="UTP-glucose glucosyltransferase"
 /protein_id="BAB02840.1"
 /db_xref="GI:11994645"
 /translation="MKILVFTSPSPALHLMATYEMAEQVLVDKDNISTVYIISFSS
 KNTSMITSLNNRLRYEIIISGDDQPTLQATDSHISGLPLRVDAVAKLVDSLDP
 APRLAGFVMDCTSMIDVANEFGVPSYFTSNAGFLGLIHTIQFMDADIDYDMSR
 LEDSDVLEVPISLTSPYPLKCLPYIFKSKEMVLPVTOARFRERKGIIVTVDDLP
 OALTFLSNGNI PRAYPGVPLIHKIKVNCVDYDKOSLTPWLDDQSPRSVYPLCRGSM
 GGFSEQYRRTALADRSGRFLMSLRASRNILRPBGETNTBEIIPGFPRTAN
 RGVYMAEQVAILAKPAIGGFVSHGNNSTLESIMFGVPAIMPLVAEQFNAPFEMV
 BELGLAVIKKMRGDLGLGRSEIVTAEIKGIIICLMEODSDVRKRVNEISEKCHVA
 LMDGSSSTALKRFIQDVTENIAMSSETS"
 complement (31809..35357,35613..35761)
 /note="gene_id:MSD21.10"
 /codon_start=1
 /evidence=not_experimental
 /product="UTP-glucose glucosyltransferase"
 /protein_id="BAB02841.1"
 /db_xref="GI:11994646"
 /translation="MGCFRGRQAREIAKALDRSSHRPLMSLRASNSIFKSKESNM
 IEEVLPESOKOTGACILPEIRRRNNMFEVFPYPTGHLRSTVEAKLVRETR
 LSIYIILPISBEGVAGSDIRIALSASNNRLREIVASDDPTETETIHKNO
 EPKASTVAKLEDYSSKPSDKIAGFLMDCTSMIDVANEFGVPSYFTSNAGFLGLIHTIQFMDADIDYDMSR
 SVYHVMCLDENVKDVSENDYADSEAVLNFPSPSRPYPKCLPHALANMMPLFVN
 QARKREMGKILVNTVALEPYLKFSSSDTPPYVPGVPLIHLNENODSKDKRLR
 IIRMDQDPSSVYPLCRGSMGFSEQYRRTALADRSGRFLMSLRASRNILRPBGETNTBEIIPGFPRTAN
 PGEFTNLEVPLEGFRTDKDIGKYGKAGVPAIMPLVAEQFNAPFEMV

Query Match 85.6%; Score 21.4; DB 15; Length 59793;
 Best Local Similarity 95.7%; Pred. No. 6.1e+02;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 AAAAAAAAACTAAGCTTGATCTT 25
 Db 45347 AAAAAAAGCTTAAGCTTGATCTT 45369

RESULT 30
 AC021453
 LOCUS Homo sapiens clone RP11-125C16, WORKING DRAFT SEQUENCE, 14
 DEFINITION
 AC021453 156165 bp DNA linear HTG 01-APR-2000
 AC021453
 VERSION Homo sapiens
 KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 1 (bases 1 to 156165)
 Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens clone RP11-125C16
 Unpublished
 2 (bases 1 to 156165)
 Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckert,J., Beda,F.,
 Boguslavsky,L., Bouckhagalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepey,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeRella,A., Dewar,K., Domino,M., Doyle,M., Fensholt,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardina,S., Grant,G., Hagoes,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kam,L., Karakas,A., Klein,J.,
 Lander,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

TITLE
 JOURNAL
 COMMENT

Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Polare,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
 Tirrell,A., Vasilev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 1, 2000 this sequence version replaced gi:6721267.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information
 Center project name: 15169
 Center clone name: 125_C_16

Summary Statistics
 Sequencing vector: MJ3; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 145142 bases at least Q40
 Consensus quality: 151009 bases at least Q40
 Consensus quality: 153345 bases at least Q20
 Insert size: 160000; agarose-fp
 Insert size: 154865; sum-of-contigs

Quality coverage: 4.1 in Q20 bases; agarose-fp
 Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1836: contig of 1836 bp in length
 * 1837 1936: gap of 100 bp
 * 1937 5796: contig of 3860 bp in length
 * 5797 5896: gap of 100 bp
 * 5897 9797: contig of 3901 bp in length
 * 9798 9897: gap of 100 bp
 * 9898 13919: contig of 4022 bp in length
 * 13920 14019: gap of 100 bp
 * 14020 17400: contig of 3381 bp in length
 * 17401 17500: gap of 100 bp
 * 17501 21253: contig of 3753 bp in length
 * 21254 21353: gap of 100 bp
 * 21354 31096: contig of 9743 bp in length
 * 31097 31196: gap of 100 bp
 * 31197 41465: contig of 10269 bp in length
 * 41466 41565: gap of 100 bp
 * 41566 50901: contig of 9336 bp in length
 * 50902 51001: gap of 100 bp
 * 51002 62041: contig of 11040 bp in length
 * 62042 62141: gap of 100 bp
 * 62142 75247: contig of 13106 bp in length
 * 75248 75347: gap of 100 bp
 * 75348 89874: contig of 14527 bp in length
 * 89875 89974: gap of 100 bp
 * 89975 116362: contig of 26388 bp in length
 * 116363 116463: gap of 100 bp
 * 116463 156165: contig of 39703 bp in length.

FEATURES
 source

1. 156165
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

```

misc_feature      /clone="RP11-125C16"
                  /clone_lib="RPCT-11 Human Male BAC"
                  1.1836
                  /note="assembly_fragment"
gap              1837..1936
                  /estimated_length=100
misc_feature      1937..5796
                  /note="assembly_fragment"
gap              5797..5896
                  /estimated_length=100
misc_feature      5897..9797
                  /note="assembly_fragment"
gap              9798..9897
                  /estimated_length=100
misc_feature      9898..13919
                  /note="assembly_fragment"
gap              13920..14019
                  /estimated_length=100
misc_feature      14020..17400
                  /note="assembly_fragment"
gap              17401..17500
                  /estimated_length=100
misc_feature      17501..21253
                  /note="assembly_fragment"
                  clone_end:SR6
                  vector_side:right"
gap              21254..21353
                  /estimated_length=100
misc_feature      21354..31096
                  /note="assembly_fragment"
gap              31097..31196
                  /estimated_length=100
misc_feature      31197..41465
                  /note="assembly_fragment"
gap              41466..41565
                  /estimated_length=100
misc_feature      41566..50901
                  /note="assembly_fragment"
gap              50902..51001
                  /estimated_length=100
misc_feature      51002..62041
                  /note="assembly_fragment"
gap              62042..62141
                  /estimated_length=100
misc_feature      62142..75247
                  /note="assembly_fragment"
gap              75248..75347
                  /estimated_length=100
misc_feature      75348..89874
                  /note="assembly_fragment"
                  vector_side:right"
gap              89875..89974
                  /estimated_length=100
misc_feature      89975..116362
                  /note="assembly_fragment"
gap              116363..116462
                  /estimated_length=100
misc_feature      116463..156165
                  /note="assembly_fragment"

```

```

Query Match      84.0%; Score 21; DB 14; Length 156165;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAAGCTTGA 21
         |||||
Db      148339 AAAAAAAAAAAGCTTGA 148359

```

```

RESULT 31
ACT00852

```

```

LOCUS      ACT00852      157980 bp      DNA      linear      PRI 29-AUG-2002
DEFINITION Homo sapiens chromosome 17, clone RP11-125C16, complete sequence.
ACCESSION  ACT00852
VERSION     ACT00852.2  GI:22539166
KEYWORDS   HTG.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homidae; Homo.
REFERENCE  1 (bases 1 to 157980)
            Birren, B., Nusbaum, C., and Lander, E.
            Homo sapiens chromosome 17, clone RP11-125C16
            Unpublished
            2 (bases 1 to 157980)
            Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
            Anderson, S., Batra, N., Bastien, V., Boguslavskiy, L., Boukhalter, B.,
            Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
            Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
            Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
            Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
            Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
            Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
            Jones, C., Kamat, A., Karatas, A., Kelle, C., Labroque, K.,
            Lamazares, R., Landers, T., Lenockzy, J., Levine, R., Liu, G.,
            Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,
            McCarthy, M., McEwan, P., McKernan, K., McNeesters, R., Meldrim, J.,
            Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
            Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
            Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
            Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
            Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
            Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
            Strause, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
            Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
            Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
            Zainoun, J., Zemdek, U., Zimmer, A. and Zody, M.
TITLE      Direct Submission
JOURNAL    Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
REFERENCE  Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS    3 (bases 1 to 157980)
            Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
            Batra, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,
            Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
            Cook, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
            Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
            Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
            Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
            Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K.,
            Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
            McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mlenga, V.,
            Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
            O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
            Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
            Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
            Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
            Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
            Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
            Zemdek, U., Zimmer, A. and Zody, M.
TITLE      Direct Submission
JOURNAL    Submitted (29-AUG-2002) Whitehead Institute/MIT Center for Genome
COMMENT    Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Aug 29, 2002 this sequence version replaced gi:17048222.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence.submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: I21587

```

```

Center clone name: 125_C_16
-----
FEATURES
Source
Location/Qualifiers
1..157980
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-125C16"
/clone_11b="RP11 Human Male BAC"
repeat_region
62..173
/rpc_family="MIR"
complement(636..832)
/rpc_family="MIR"
repeat_region
833..888
/rpc_family="MER3"
complement(891..1207)
/rpc_family="AluDb"
repeat_region
1208..1283
/rpc_family="MER3"
1232..1236
/note="< 30 qual SINGL region"
complement(1303..2370)
/rpc_family="L1PA4"
1330..1354
/note="<30 qual SINGL region"
2371..3213
/rpc_family="L1PA4"
complement(3266..3408)
/rpc_family="Charlie7"
3506..3538
/rpc_family="L2"
3539..3653
/rpc_family="MER47A"
complement(3654..3926)
/rpc_family="AluSc"
3927..4141
/rpc_family="MER47A"
4142..4438
/rpc_family="AluY"
4439..4473
/rpc_family="MER47A"
4474..4561
/rpc_family="L2"
complement(4601..4700)
/rpc_family="L2"
repeat_region
complement(4708..4845)
/rpc_family="MER5A"
repeat_region
complement(4944..5104)
/rpc_family="MIR"
complement(5112..5532)
/rpc_family="L2"
complement(5546..5822)
/rpc_family="AluDb"
complement(5823..5999)
/rpc_family="L2"
6000..6216
/rpc_family="AluDb"
complement(6217..6292)
/rpc_family="L2"
6447..6752
/rpc_family="AluDb"
7412..7472
/rpc_family="A-rich"
7789..7907
/rpc_family="A-rich"
8642..8683
/rpc_family="r-rich"
8934..8978
/rpc_family="CCA)n"
9084..9123
/rpc_family="(CAG)n"

```

```

repeat_region 10990..11017
/rpc_family="AT_rich"
repeat_region 11225..11326
/rpc_family="L2"
repeat_region complement(11372..11504)
/rpc_family="MLT1B"
repeat_region 13034..13209
/rpc_family="MIR"
complement(15480..15590)
/rpc_family="L2"
repeat_region 15703..15988
/rpc_family="AluSc"
complement(16005..16318)
/rpc_family="L2"
16412..16555
/rpc_family="MER106B"
16825..17099
/rpc_family="AluDb"
17241..17261
/rpc_family="AT_rich"
17280..17310
/rpc_family="(TTTG)n"
complement(17366..17658)
/rpc_family="AluSc"
17707..17808
/rpc_family="MIR"
18250..18277
/rpc_family="AT_rich"
complement(18576..18864)
/rpc_family="MER3"
complement(18926..19077)
/rpc_family="AluDb"
20786..20952
/rpc_family="AluDb"
20962..21087
/rpc_family="MIR"
complement(21088..21390)
/rpc_family="AluSc"
complement(21419..21646)
/rpc_family="AluDb"
21647..21715
/rpc_family="MIR"
complement(21969..22223)
/rpc_family="L1ME4A"
complement(22288..22411)
/rpc_family="MIR3"
complement(22950..23256)
/rpc_family="AluSc"
24607..24789
/rpc_family="FAM"
24846..25144
/rpc_family="AluDb"
25229..25254
/rpc_family="AT_rich"

```

```

Query Match 84.0%; Score 21; DB 8; Length 157980;
Best Local Similarity 100.0%; Pred. No. 5; 9e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAAGCTGA 21
Db 110408 AAAAAAAAACTAAAGCTGA 110428

RESULT 32
LOCUS BV267418 654 bp DNA linear STS 22-JAN-2005
DEFINITION S235F6522RG11.T0 ItalianGreyhound Canis familiaris STS genomic.
Sequence tagged site.
ACCESSION BV267418
VERSION BV267418.1 GI:57329960
KEYWORDS STS.
SOURCE Canis familiaris (dog)

```

ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
REFERENCE 1 (bases 1 to 654)
TITLE Lindblad-Toh, K.
JOURNAL The genome sequence of Canis familiaris
COMMENT Unpublished (2004)

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersti@genome.wi.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 654

Protocol:
WGS-discovery (WGS):
Paired-end low-coverage whole genome shotgun reads were generated
from 9 breeds
(German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
Retriever, English
Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
Water Dog -100,000 each)
and five other canids (Chinese, Alaskan, Indian and Spanish Gray
Wolf as well as the
Californian Coyote).
The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
and SNP detection was
carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
485941 SNPs were
annotated with alleles from the boxer and the breed or canid from
which the particular
read came. The validation rate for these SNPs was estimated at
approximately 98%.
WGA-discovery (WGA) of Boxer/Poodle SNPs:
A second set of SNPs was generated using a similar methodology
except that the contigs
from the 1.5x poodle assembly (Kirkness 2003) were used instead of
WGS reads. Since this
sequence lacked base quality scores, arbitrary quality scores of
phred 40 were assigned
before the poodle sequence was placed uniquely on the CanFam1.0
boxer assembly and SNP
detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
with alleles from the
boxer and the poodle. The validation rate for these SNPs was
estimated at approximately TBD%.
Internal-WGA-discovery (I-WGA):
A third set of SNPs were discovered by comparing reads in the WGA
assembly. SNPs were
defined as mismatch positions that had a base quality of >= 30 on
both reads in a region.
that aligned without gaps, and with at most one additional mismatch
in the ten flanking
bases. For each allele, at least one additional read had to confirm
it. 731476 SNPs were
annotated with alleles between the two boxer alleles. The
validation rate for these SNPs
was estimated at approximately TBD%.

FEATURES
source
1..654
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="ItalianGreyhound"
/db_xref="taxon:9615"
/map="19 22-611 16431893-16432486"
/clone_lib="ItalianGreyhound"
<1..>654

STS
ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 654;
Best Local Similarity 91.7%; Pred. No. 4.9e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTAAGCTTGATCT 24
|||||
Db 404 AAAAAAAAAAAGAAAGCTTGATCT 427

RESULT 33
LOCUS CO730691 4322 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 16625 from Patent WO02068579.
ACCESSION CO730691
VERSION CO730691.1 GI:42305224
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.

REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 16625 06-SEP-2002;
PE Corporation (NY) (US)

FEATURES
source
1..4322
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 4322;
Best Local Similarity 91.7%; Pred. No. 2.5e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTAAGCTTGATCT 24
|||||
Db 986 AAAAAAAAAAAGAAAGCTTGATCT 1009

RESULT 34
LOCUS AX251316 13606 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 284 from Patent WO0168912.
ACCESSION AX251316
VERSION AX251316.1 GI:15984739
KEYWORDS
SOURCE
ORGANISM synthetic construct
synthetic construct
other sequences; artificial sequences.

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 284 20-SEP-2001;
Epigenomics AG (DE)

FEATURES
source
1..13606
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:35630"
/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
    |||||
Db 5439 AAAAAAAAACTAAAGCTTAATCT 5416

RESULT 35
AX278003/c 13606 bp DNA linear PAT 01-NOV-2001
LOCUS Sequence 166 from Patent WO0177375.
DEFINITION AX278003
ACCESSION AX278003
VERSION AX278003.1 GI:16605075
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 Olek,A., Piepenbrock,C. and Berlin,K.
  Diagnosis of diseases associated with gene regulation
  Patent: WO 0177375-A 166 18-OCT-2001;
  Epigenomics AG (DE)
FEATURES
    source
    Location/Qualifiers
        1..13606
        /organism="synthetic construct"
        /mol_type="unassigned DNA"
        /db_xref="taxon:32630"
        /note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
    |||||
Db 5439 AAAAAAAAACTAAAGCTTAATCT 5416

RESULT 36
AX323700/c 13606 bp DNA linear PAT 07-JAN-2002
LOCUS Sequence 188 from Patent WO0192565.
DEFINITION AX323700
ACCESSION AX323700
VERSION AX323700.1 GI:18094448
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 Olek,A., Piepenbrock,C. and Berlin,K.
  Diagnosis of diseases associated with dna transcription
  Patent: WO 0192565-A 188 06-DEC-2001;
  Epigenomics AG (DE)
FEATURES
    source
    Location/Qualifiers
        1..13606
        /organism="synthetic construct"
        /mol_type="unassigned DNA"
        /db_xref="taxon:32630"
        /note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
    |||||
Db 5439 AAAAAAAAACTAAAGCTTAATCT 5416

RESULT 37
AX346713/c 13606 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 1784 from Patent WO0200928.
DEFINITION

```

```

ACCESSION AX346713
VERSION AX346713.1 GI:18494599
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 Olek,A., Piepenbrock,C. and Berlin,K.
  Diagnosis of diseases associated with the immune system
  Patent: WO 0200928-A 1784 03-JAN-2002;
  Epigenomics AG (DE)
FEATURES
    source
    Location/Qualifiers
        1..13606
        /organism="synthetic construct"
        /mol_type="unassigned DNA"
        /db_xref="taxon:32630"
        /note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 1.7e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
    |||||
Db 5439 AAAAAAAAACTAAAGCTTAATCT 5416

RESULT 38
AL136095 50133 bp DNA linear PRI 18-MAY-2005
LOCUS Human DNA sequence from clone RPL-25L16 on chromosome 6, complete
sequence.
DEFINITION AL136095
ACCESSION AL136095
VERSION AL136095.25 GI:13234794
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.
1 (bases 1 to 50133)
REFERENCE
AUTHORS Martin,S
JOURNAL Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Mar 5, 2001 this sequence version replaced GI:13184276.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Bm, BMB; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Ch6
RPL-25L16 is from the library RPLC1-1 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,

```



```

* 23569 23668: gap of 100 bp
* 23669 24408: contig of 740 bp in length
* 24409 24508: gap of 100 bp
* 24509 25255: contig of 747 bp in length
* 25256 25355: gap of 100 bp
* 25356 26112: contig of 757 bp in length
* 26113 26212: gap of 100 bp
* 26213 26961: contig of 749 bp in length
* 26962 27061: gap of 100 bp
* 27062 27791: contig of 730 bp in length
* 27792 27891: gap of 100 bp
* 27892 28646: contig of 755 bp in length
* 28647 28746: gap of 100 bp
* 28747 29467: contig of 721 bp in length
* 29468 29567: gap of 100 bp
* 29568 30295: contig of 728 bp in length
* 30296 30395: gap of 100 bp
* 30396 31141: contig of 746 bp in length
* 31142 31241: gap of 100 bp
* 31242 31979: contig of 738 bp in length
* 31980 32079: gap of 100 bp
* 32080 32804: contig of 725 bp in length
* 32805 32904: gap of 100 bp
* 32905 33641: contig of 737 bp in length
* 33642 33741: gap of 100 bp
* 33742 34455: contig of 714 bp in length
* 34456 34555: gap of 100 bp
* 34556 35382: contig of 827 bp in length
* 35383 35482: gap of 100 bp
* 35483 36215: contig of 733 bp in length
* 36216 37069: contig of 754 bp in length
* 37070 37169: gap of 100 bp
* 37170 37922: contig of 753 bp in length
* 37923 38022: gap of 100 bp
* 38023 38765: contig of 743 bp in length
* 38766 38865: gap of 100 bp
* 38866 39605: contig of 740 bp in length
* 39606 39705: gap of 100 bp
* 39706 40450: contig of 745 bp in length
* 40451 40550: gap of 100 bp
* 40551 41299: contig of 749 bp in length
* 41300 41399: gap of 100 bp
* 41400 42174: contig of 775 bp in length
* 42175 42274: gap of 100 bp
* 42275 43020: contig of 746 bp in length
* 43021 43120: gap of 100 bp
* 43121 43875: contig of 755 bp in length
* 43876 43975: gap of 100 bp
* 43976 44740: contig of 765 bp in length
* 44741 44840: gap of 100 bp
* 44841 45568: contig of 728 bp in length
* 45569 45668: gap of 100 bp
* 45669 46404: contig of 736 bp in length
* 46405 46504: gap of 100 bp
* 46505 47267: contig of 763 bp in length
* 47268 47367: gap of 100 bp
* 47368 48079: contig of 712 bp in length
* 48080 48179: gap of 100 bp
* 48180 48914: contig of 735 bp in length
* 48915 49014: gap of 100 bp
* 49015 49761: contig of 747 bp in length
* 49762 49861: gap of 100 bp
* 49862 50615: contig of 754 bp in length
* 50616 50715: gap of 100 bp
* 50716 51449: contig of 734 bp in length
* 51450 51549: gap of 100 bp
* 51550 52285: contig of 736 bp in length
* 52286 52385: gap of 100 bp
* 52386 53131: contig of 746 bp in length
* 53132 53231: gap of 100 bp
* 53232 53985: contig of 754 bp in length
* 53986 54085: gap of 100 bp

```

```

* 54086 54821: contig of 736 bp in length
* 54822 54921: gap of 100 bp
* 54922 55672: contig of 751 bp in length
* 55673 55772: gap of 100 bp
* 55773 56536: contig of 764 bp in length
* 56537 56637: gap of 100 bp
* 56637 57352: contig of 716 bp in length
* 57353 57452: gap of 100 bp

Query Match      83.2%; Score 20.8; DB 14; Length 61569;
Best Local Similarity 91.7%; Pred. No. 9.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAACTTAAGCTGATCT 24
Db 59164 AAAAAAAAAAACAACTGATCT 59141

RESULT 40
AP008123
LOCUS
DEFINITION
AP008123
LOCUS corniculatus var. japonicus chromosome 4 clone Lj11c16, ***
SEQUENCING IN PROGRESS ***, 10 unordered pieces.
AP008123
VERSION
AP008123.1 GI:56806430
KEYWORDS
HTG; HTGS PHASE1.
SOURCE
Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM
Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
REFERENCE
1
Kaneke,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
Structural Analysis of a Lotus japonicus Genome. XI. Sequence
Features and Mapping of Nine hundred twenty-one TAC Clones
Unpublished
2 (bases 1 to 80115)
Sato,S.
Direct Submission
Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu,
Chiba, 292-0818, Japan (E-mail: ssato@kazusa.or.jp,
Url: http://www.kazusa.or.jp/, Tel: 81-438-52-3935 (ex. 2337),
Fax: 81-438-52-3934).
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
10254: contig of 10254 bp in length
* 10255 10354: gap of unknown length
* 10355 11321: contig of 967 bp in length
* 11322 11421: gap of unknown length
* 11422 12425: contig of 1004 bp in length
* 12426 12525: gap of unknown length
* 12526 13449: contig of 924 bp in length
* 13450 13549: gap of unknown length
* 13550 15466: contig of 1937 bp in length
* 15467 15586: gap of unknown length
* 15587 17828: contig of 2242 bp in length
* 17829 17928: gap of unknown length
* 17929 21342: contig of 3414 bp in length
* 21343 21442: gap of unknown length
* 21443 29197: contig of 7755 bp in length
* 29198 29297: gap of unknown length
* 29298 41474: contig of 12177 bp in length
* 41475 80115: contig of 38541 bp in length.
* 41575
Location/Qualifiers

```

FEATURES


```

source
1. 8015
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"
/chromosome="4"
/clone="UJ11C16"
/clone_lib="UJ1 library"
/notes="TAC clone: TM1509, synonym: Lotus japonicus"
gap
10255. 10354
/estimated_length=unknown
11322. 11421
/estimated_length=unknown
12426. 12525
/estimated_length=unknown
13450. 13549
/estimated_length=unknown
15487. 15586
/estimated_length=unknown
17829. 17928
/estimated_length=unknown
21343. 21442
/estimated_length=unknown
29198. 29297
/estimated_length=unknown
41475. 41574
/estimated_length=unknown

ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 80115;
Best Local Similarity 91.7%; Pred. No. 8.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTGATCT 24
Db 22732 AAAAAAAAACTAAGCTGTTT 22755

RESULT 41
LOCUS AC021876 104668 bp DNA linear PRI 04-JUN-2002
DEFINITION Homo sapiens BAC clone GSI-117B4 from 7, complete sequence.
ACCESSION AC021876
VERSION AC021876.5 GI:13446341
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 104668)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
9847074
2 (bases 1 to 104668)
AUTHORS Du, F., Stoneking, T. and Moeller, D.
TITLE The sequence of Homo sapiens BAC clone GSI-117B4
JOURNAL Unpublished (2001)
3 (bases 1 to 104668)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 104668)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 104668)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 104668)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 104668)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
8 (bases 1 to 104668)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
9 (bases 1 to 104668)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 25, 2001 this sequence version replaced gi:9665195.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_GSI117B04
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc.
(http://www.genomesystems.com).
Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelobAC
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTA-271G13, 200 bp overlap; the
clone sequenced to the right is AC023375. Actual start of this
clone is at base position 102112 of CTA-271G13; actual end is at
base position 104668 of GSI-117B4.

The sequence from 80185 to 80225 was derived from PCR product of

```

FEATURES
 source

GSI-117B4 BAC DNA.
 Location/Qualifiers
 1..104668
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"
 /clone_id="GSI-117B4"
 /clone_lib="GSBAC1"

repeat_region
 1042..1139
 /rpt_family="Alu"

misc_feature
 1733..2070
 /note="similar to EST BF469008 (NID:g11538191)"

repeat_region
 2062..2084
 /rpt_family="(A)n"

repeat_region
 2277..2563
 /rpt_family="Alu"

misc_feature
 2461..2908
 /note="similar to EST R99288 (NID:g985889) Yq71b08.b1"

repeat_region
 3168..3193
 /rpt_family="AT-rich"

repeat_region
 3432..3453
 /rpt_family="AT-rich"

misc_feature
 3447..3677
 /note="similar to EST A1536708 (NID:g4450843) tm87g10.x1"

repeat_region
 3487..3593
 /rpt_family="Alu"

repeat_region
 3654..3956
 /rpt_family="Alu"

repeat_region
 4476..4694
 /rpt_family="MALR"

repeat_region
 4695..4821
 /rpt_family="MALR"

repeat_region
 4822..5056
 /rpt_family="Alu"

repeat_region
 5363..5487
 /rpt_family="MALR"

repeat_region
 5620..5728
 /rpt_family="Alu"

repeat_region
 6641..6750
 /rpt_family="MERL-type"

repeat_region
 6966..7147
 /rpt_family="MIR"

repeat_region
 7826..8136
 /rpt_family="MIR"

repeat_region
 8384..8663
 /rpt_family="Alu"

repeat_region
 8596..8944
 /rpt_family="Alu"

misc_feature
 8673..8760
 /note="similar to EST AX070710 (NID:g12580495)"

repeat_region
 8769..8789
 /rpt_family="L1"

repeat_region
 9047..9305
 /rpt_family="AT-rich"

misc_feature
 9047..9280
 /note="similar to EST BE308143 (NID:g9164347)"

misc_feature
 9052..9177
 /note="similar to EST A1606545 (NID:g4615712) vb29b12.y1"

misc_feature
 9052..9177
 /note="similar to EST A163596 (NID:g1739542) mm38g02.r1"

misc_feature
 9052..9177
 /note="similar to EST AU127673 (NID:g10988027)"

misc_feature
 9826..10088
 /note="similar to EST BF241103 (NID:g1155028)"

repeat_region
 10215..10372
 /rpt_family="MIR"

repeat_region
 10437..10549
 /rpt_family="MIR"

misc_feature
 10706..10841
 /note="similar to EST AU127673 (NID:g10988027)"

misc_feature
 10706..10841

misc_feature
 10706..10740
 /note="similar to EST BF241103 (NID:g1155028)"

misc_feature
 10723..10841
 /note="similar to EST A163596 (NID:g1739542) mm38g02.r1"

repeat_region
 12253..12638
 /note="similar to EST AA474176 (NID:g2202403) ves2g08.r1"

misc_feature
 12928..13050
 /rpt_family="MALR"

misc_feature
 12928..13050
 /note="similar to EST AA474176 (NID:g2202403) ves2g08.r1"

misc_feature
 12928..13050
 /note="similar to EST AU127673 (NID:g10988027)"

misc_feature
 12928..13050
 /note="similar to EST BF217988 (NID:g1111574)"

misc_feature
 12945..13050
 /note="similar to EST BF241103 (NID:g1155028)"

repeat_region
 13115..13172
 /note="similar to EST AV612377 (NID:g9748047)"

repeat_region
 13571..13877
 /rpt_family="L2"

repeat_region
 13571..13877
 /rpt_family="Alu"

Query Match 83.2%; Score 20.8; DB 8; Length 104668;
 Best Local Similarity 91.7%; Pred. No. 8e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAACTAAGCTGATCT 24
 Db 29262 AAAAAAAAAAACTAAGTTTATCT 29285

RESULT 42
 LOCUS AP001576 105199 bp DNA linear PRI 16-AUG-2001
 DEFINITION Homo sapiens genomic DNA, chromosome 6q25.2, clone:KB1F5.
 ACCESSION AP001576
 VERSION AP001576.3 GI:15208253
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.

REFERENCE 1
 AUTHORS Shimizu,N. and Asakawa,S.
 TITLE Homo sapiens DNA chromosome 6 SEQUENCE
 JOURNAL Published Only in Database (2000)
 REFERENCE 2 (bases 1 to 105199)
 AUTHORS Shimizu,N. and Asakawa,S.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2000) Nobuyoshi Shimizu, Keio University, School
 of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-Ku, Tokyo
 160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
 Tel:81-3-3351-2370, Fax:81-3-3351-2370)
 On Aug 16, 2001 this sequence version replaced gi:8096256.

COMMENT
 FEATURES
 location/Qualifiers
 1..105199
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6q25.2"
 /clone="KB1F5"
 /cell_line="FUEB 14 - 14"
 /clone_lib="Keio BAC library"

repeat_region
 166..474
 /evidence=not experimental
 /rpt_family="AluSx"

repeat_region
 520..590
 /evidence=not experimental
 /rpt_family="AT-rich"

repeat_region
 complement(912..1017)
 /evidence=not experimental

```
repeat_region /rpt_family="FLAM_C"
1199. .1490 /evidence=not_experimental
/rpt_family="MLT11"
repeat_region 1844. .1922 /evidence=not_experimental
/rpt_family="TATATG)n"
repeat_region 2732. .3028 /evidence=not_experimental
/rpt_family="AluSx"
repeat_region complement(3323. .3404)
/evidence=not_experimental
/rpt_family="L2"
repeat_region complement(3608. .3816)
/evidence=not_experimental
/rpt_family="L2"
repeat_region complement(4573. .4720)
/evidence=not_experimental
/rpt_family="MIR"
repeat_region complement(7641. .8105)
/evidence=not_experimental
/rpt_family="L1PA16"
repeat_region 8106. .8852 /evidence=not_experimental
/rpt_family="L1PA16"
repeat_region 8855. .8884 /evidence=not_experimental
/rpt_family="At_rich"
repeat_region 9163. .9192 /evidence=not_experimental
/rpt_family="At_rich"
repeat_region complement(10418. .10545)
/evidence=not_experimental
/rpt_family="AluSx"
repeat_region 10546. .10585 /evidence=not_experimental
/rpt_family="MER58c"
repeat_region complement(10797. .10797)
/evidence=not_experimental
/rpt_family="Alu"
repeat_region complement(10837. .11091)
/evidence=not_experimental
/rpt_family="AluSc"
repeat_region 11771. .11813 /evidence=not_experimental
/rpt_family="L2"
repeat_region complement(12629. .12706)
/evidence=not_experimental
/rpt_family="MIR"
repeat_region 12903. .13167 /evidence=not_experimental
/rpt_family="AluJo"
repeat_region 13173. .13453 /evidence=not_experimental
/rpt_family="AluSg"
repeat_region 13871. .13920 /evidence=not_experimental
/rpt_family="At_rich"
repeat_region complement(14938. .15193)
/evidence=not_experimental
/rpt_family="AluSg"
repeat_region 15379. .15453 /evidence=not_experimental
/rpt_family="L2"
repeat_region complement(15501. .15754)
/evidence=not_experimental
/rpt_family="LTR37B"

repeat_region complement(16177. .16924)
/evidence=not_experimental
/rpt_family="MER31-internal"
repeat_region complement(17232. .17377)
/evidence=not_experimental
/rpt_family="MER31-internal"
repeat_region 18854. .18955 /evidence=not_experimental
/rpt_family="At_rich"
repeat_region 19314. .19624 /evidence=not_experimental
/rpt_family="AluSx"
repeat_region 20973. .21071 /evidence=not_experimental
/rpt_family="MIR"
repeat_region complement(22061. .22369)
/evidence=not_experimental
/rpt_family="AluSx"
repeat_region 22428. .22463 /evidence=not_experimental
/rpt_family="At_rich"
repeat_region complement(22469. .22781)
/evidence=not_experimental
/rpt_family="AluJo"
repeat_region complement(22820. .23154)
/evidence=not_experimental
/rpt_family="MER7A"
repeat_region 23362. .23397 /evidence=not_experimental
/rpt_family="At_rich"
repeat_region 23468. .23503 /evidence=not_experimental
/rpt_family="At_rich"
repeat_region 24148. .24175 /evidence=not_experimental
/rpt_family="TA)n"
repeat_region complement(24576. .24879)
/evidence=not_experimental
/rpt_family="AluSx"
repeat_region 24980. .25344 /evidence=not_experimental
/rpt_family="MLT2F"
repeat_region complement(25365. .25422)
/evidence=not_experimental
/rpt_family="MIR"
repeat_region complement(25372. .25487)
/evidence=not_experimental
/rpt_family="L2"
repeat_region 25515. .25637 /evidence=not_experimental
/rpt_family="MER94"
repeat_region 25822. .25868 /evidence=not_experimental
/rpt_family="Alu"
repeat_region 26514. .26818 /evidence=not_experimental
/rpt_family="AluVa5"
repeat_region 27318. .27344 /evidence=not_experimental
/rpt_family="TTTTG)n"
repeat_region 27983. .28027 /evidence=not_experimental
/rpt_family="AT187. .28382)
complement(28187. .28382)
/evidence=not_experimental
/rpt_family="MER20"
repeat_region 28857. .28993 /evidence=not_experimental
/rpt_family="FLAM_C"
repeat_region 29511. .29565 /evidence=not_experimental
/rpt_family="TG)n"
repeat_region 29582. .29620
```

[illegible]

TAYLOR, T., TELFORD, B., THOMAS, N., THOMAS, S., UEMANI, K., VAQUEZ, L.,
 VERA, V., VILLALON, D., VINSON, R., WANG, Q., WANG, S., WARD-MOORE, S.,
 WARREN, R., WASHINGTON, C., WATLINGTON, S., WILLIAMS, G.,
 WILLIAMSON, A., WLECZYK, R., WOODEN, S., WORLEY, K., WU, C., WU, Y.,
 WU, Y. F., ZHOU, J., ZORRILLA, S., NAYLOR, S. L., WEINSTECK, G. and
 Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 130669)
 Worley, K.C.
 Direct Submission
 Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 130669)
 Worley, K.C.
 Direct Submission
 Submitted (29-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 130669)
 Worley, K.C.
 Direct Submission
 Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 130669)
 Worley, K.C.
 Direct Submission
 Submitted (25-FEB-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Aug 30, 2002 this sequence version replaced gi:22538686.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:
 STS are identified using ePCR (genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL:
<http://www.hgsc.bcm.tmc.edu:8086/quality.info/genbank.annotation.htm>.

ml.

Location/Qualifiers
 1..130669
 /organism="Homo sapiens"

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAACCAAGCTTGATCT 24
 Db 4121 AAAAAAAAAACCAAGCTTGATCT 4144

RESULT 45
 AC105752 142525 bp DNA linear PRI 30-APR-2002
 LOCUS Homo sapiens chromosome 3 clone RP11-45K8, complete sequence.
 AC105752 AC015627
 AC105752.2 GI:20340480
 HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.

REFERENCE
 AUTHORS 1 (bases 1 to 142525)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
 Haugen, E.D.

TITLE
 JOURNAL Direct Submission
 REFERENCE Unpublished
 AUTHORS 2 (bases 1 to 142525)
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 JOURNAL Direct Submission
 REFERENCE Submitted (09-JAN-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 142525)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
 Haugen, E.D.

TITLE
 JOURNAL Direct Submission
 REFERENCE Submitted (30-APR-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Apr 30, 2002 this sequence version replaced gi:18093013.

COMMENT
 ----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: http://www.genome.washington.edu
 Contact: uwgchgs@u.washington.edu
 Drafting Center: BCM
 ----- Project Information
 Center project name: chr-3
 Center clone name: RP11-45K8 (bc0612)

----- Summary Statistics
 Sequencing vector: plasmid; 100% of reads
 Chemistry: Dye-terminator ET; 92% of reads
 Chemistry: Dye-terminator Big Dye; 8% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 142405 bases at least Q40
 Consensus quality: 142509 bases at least Q30
 Consensus quality: 142525 bases at least Q20
 Insert size: 142525; sum-of-contigs
 Quality coverage: 8.5x in Q20 bases; sum-of-contigs

Overlapping Sequences:
 3: Mapping in progress
 5: AP000497

----- Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:

all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., Phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
 fingerprinting. Comparison of the experimentally derived digest
 fragments with sequence-predicted fragments is given below.
 The electronically-digested sequence consists of both insert and
 vector, in order to accurately represent the entire circular BAC.
 Small fragments below a variable cutoff (approximately 400-800 bp)
 are not resolved in the fingerprint and hence do not appear
 in the table. There are no significant remaining discrepancies
 between the experimental and predicted values. Uniquely ordered
 fragments are separated by dashed lines.

ECORI	BglII	NsiI
SeqDerMap Fingerprint	SeqDerMap Fingerprint	SeqDerMap Fingerprint
8696	8768	3424
6	<850	2067
3369	3244	9003
2152	2119	2693
3210	3244	6108
2363	2328	5921
7823	7718	747
1696	1650	4836
188	<850	4739
142	<850	1482
1798	1785	2614
2902	2868	3528
1447	1420	653
1023	1004	4115
1479	1420	6483
1927	1896	2349
3589	3585	2687
2289	2328	3258
1343	1420	4996
884	891	3170
4921	4877	168
253	<850	12591
708	<850	5758
573	<850	3805
5029	4877	10858
		10797
		2795
		2864

8084	8069	10123	10001	1358	1351
489	<850	4473	4659	982	983
1915	1896	12019	12033	1607	1601
3580	3585	9063	9035	2529	2438
4775	4649	7493	7539	2683	2687
1823	1785			592	<850
466	<850			57	<850
13054	13253			909	867
769	<850			1357	1351
3126	3244			4804	4827
781	793			4122	4088
140	<850			7122	7138
2284	2328			11498	11487
1819	1785			853	867
3611	3585			4058	4088
4696	4649			3849	3852
2952	2975			2203	2169
13444	13253			3743	3719
6218	6247			2708	2687
11501	11505			22	<850
5927	5975			11475	11487
				222	<850
				10097	10220

FEATURES

source 1.142525 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-45K8"
 /clone_lib="RPCT human BAC library 11"

Query Match 83.2%; Score 20.8; DB 8; Length 142525;
 Best Local Similarity 91.7%; Pred. No. 7.2e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTTAACTTGATCT 24
 Db 116659 AAAAAAAAAAAAGCTTGATCT 116682

RESULT 46
 AP000497 146596 bp DNA linear PRI 28-SEP-1999
 LOCUS Homo sapiens genomic DNA, chromosome 3p21.3, clone:301 to 308,
 DEFINITION anti-oncogene region, section 5/5.
 ACCESSION AP000497
 VERSION AP000497.1 GI:5926684

KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.

REFERENCE
 AUTHORS Nakamura,Y., Iwamura,M., Daigo,Y., Tamari,M. and Ishikawa,S.
 TITLE DNA sequence analysis of a 1.2-Mb region on chromosome 3p21.3
 JOURNAL Published Only in Database (1999)
 REFERENCE 2 (bases 1 to 146596)

AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
 TITLE Direct Submission
 JOURNAL Submitted (21-SEP-1999) Mika Hirakawa, Japan Science and Technology Corporation (JST), Advanced Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mka@tokyo.jst.go.jp, URL:http://www-alls.tokyo.jst.go.jp/, Tel:81-3-5214-8491, Fax:81-3-5214-8470)

COMMENT

This sequence is conducted by Japanese Foundation for Cancer Research as a JST sequencing Team.
 Principal Investigator: Yusuke Nakamura Ph.D
 Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,
 Yusuke@jgc.ims.u-tokyo.ac.jp
 The sequence is submitted by Human Genome Sequencing in ALIS project of JST
 Japan Science and Technology Corporation (JST)
 5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
 For further information about this sequences, please visit our
 sequence archive Web site (http://www-alls.tokyo.jst.go.jp/HGS/cep.html) or send email to webmaster@www-alls.tokyo.jst.go.jp
 This entry is upstream sequence of AP000498.1 with a gap between them.

FEATURES

source

1.146596 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3p21.3"
 /note="301-308"
 complement(38091..38252)
 /note="sts88510:The location is between each flanking site of PCR primers."
 /db_xref="GDB:4569367"
 complement(38233..38336)
 /note="SHGC-33953:The location is between each flanking site of PCR primers."
 /db_xref="GDB:6456343"
 complement(44376..44577)
 /standard_name="D3S307"
 /note="WI-4833:The location is between each flanking site of PCR primers."
 /db_xref="GDB:458448"
 complement(58686..58822)
 /note="SHGC-17375:The location is between each flanking site of PCR primers."
 /db_xref="GDB:6455961"
 72032..82068
 /gene="Hs.103665"
 /note="match Hs.103665 (D88154) with account of intervene sequences."
 complement(82404..104487)
 /gene="PLCD1"
 complement(82404..104487)
 /gene="PLCD1"
 /note="match PLCD1 (U09117) with account of intervene sequences."
 /db_xref="GDB:6075994"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 146596;

```

Beat Local Similarity 91.7%; Pred. No. 7.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAAAGCTTGATCT 24
Db 56065 AAAAAAAAAAAGCTTGATCT 56088

RESULT 47
AP004289/c 151991 bp DNA linear PRI 04-DEC-2001
LOCUS Homo sapiens genomic DNA, chromosome 8q23, clone: KB1844F2.
DEFINITION AP004289
ACCESSION AP004289
VERSION AP004289.2 GI:17299201
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE 1
AUTHORS Shimizu, N. and Asakawa, S.
TITLE Homo sapiens DNA chromosome 8 SEQUENCE
JOURNAL Published Only in Database (2001)
REFERENCE 2 (bases 1 to 151991)
AUTHORS Shimizu, N. and Asakawa, S.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2001) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail: nshimizu@cmb.med.keio.ac.jp,
Tel:81-3-3351-2370, Fax:81-3-3351-2370)
On Dec 3, 2001 this sequence version replaced gi:16303298.

COMMENT
FEATURES
            source
                location/Qualifiers
                    1..151991
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="8"
                        /map="8q23"
                        /clone="KB1844F2"
                        /cell_line="FLEB 14 - 14"
                        /cell_type="pre-pro-B cell"
                        /clone_11b="Keio BAC library"
                        1540..1769
                            /evidence=not experimental
                            /rpt_family="MIR"
                        1768..1888
                            /evidence=not experimental
                            /rpt_family="L2"
                        complement(2007..2270)
                            /evidence=not experimental
                            /rpt_family="ALuOb"
                        2641..2780
                            /evidence=not experimental
                            /rpt_family="MERSB"
                        2789..2998
                            /evidence=not experimental
                            /rpt_family="MERS6A"
                        3021..3297
                            /evidence=not experimental
                            /rpt_family="ALuSg"
                        3398..3526
                            /evidence=not experimental
                            /rpt_family="F2"
                        5235..5276
                            /evidence=not experimental
                            /rpt_family="(TCC)n"
                        complement(5327..5854)
                            /evidence=not experimental
                            /rpt_family="LIMB5"
                        complement(5855..6149)
                            /evidence=not experimental
                            /rpt_family="ALuYb8"
                    repeat_region complement(6150..6182)
                        /evidence=not experimental
                        /rpt_family="LIMB5"
                    repeat_region 6187..6335
                        /evidence=not experimental
                        /rpt_family="FRAM"
                    repeat_region complement(6358..7093)
                        /evidence=not experimental
                        /rpt_family="LIMB5"
                    repeat_region complement(7100..7255)
                        /evidence=not experimental
                        /rpt_family="FRAM"
                    repeat_region complement(7263..9246)
                        /evidence=not experimental
                        /rpt_family="LIMB6"
                    repeat_region complement(9247..9968)
                        /evidence=not experimental
                        /rpt_family="LIMB6"
                    repeat_region complement(9969..10423)
                        /evidence=not experimental
                        /rpt_family="LIMB6"
                    repeat_region complement(10676..11048)
                        /evidence=not experimental
                        /rpt_family="LIMD"
                    repeat_region 11049..11075
                        /evidence=not experimental
                        /rpt_family="(TTTA)n"
                    repeat_region complement(11077..11353)
                        /evidence=not experimental
                        /rpt_family="ALuBx"
                    repeat_region complement(11354..12664)
                        /evidence=not experimental
                        /rpt_family="LIMD"
                    repeat_region 12664..13251
                        /evidence=not experimental
                        /rpt_family="LIM4"
                    repeat_region 13251..13388
                        /evidence=not experimental
                        /rpt_family="L2"
                    repeat_region complement(13581..13636)
                        /evidence=not experimental
                        /rpt_family="MSTC"
                    repeat_region 13636..14012
                        /evidence=not experimental
                        /rpt_family="(CA)n"
                    repeat_region 14012..14522
                        /evidence=not experimental
                        /rpt_family="AT rich"
                    repeat_region complement(14523..15591)
                        /evidence=not experimental
                        /rpt_family="MIR"
                    repeat_region 15591..16410
                        /evidence=not experimental
                        /rpt_family="LIMC"
                    repeat_region complement(16411..18164)
                        /evidence=not experimental
                        /rpt_family="LIMC1"
                    repeat_region 18164..19197
                        /evidence=not experimental
                        /rpt_family="(TTG)n"
                    repeat_region complement(19198..19487)
                        /evidence=not experimental
                        /rpt_family="ALuOb"
                    repeat_region 19487..19682
                        /evidence=not experimental
                        /rpt_family="HAL1"
                    repeat_region complement(19683..20283)
                        /evidence=not experimental
                        /rpt_family="MULTID"
                    repeat_region 20283..21474
                        /evidence=not experimental
                        /rpt_family="HAL1"
                    repeat_region 21474..22650
                        /evidence=not experimental
                        /rpt_family="HAL1"

```



```

/evidence=not experimental
repeat_region /rpt_family="Alu0"
22933. .23094 /evidence=not experimental
repeat_region /rpt_family="THE1C"
23080. .23362 /evidence=not experimental
repeat_region /rpt_family="THE1C"
23363. .23874 /evidence=not experimental
repeat_region /rpt_family="THE1C-internal"
23875. .24182 /evidence=not experimental
repeat_region /rpt_family="AluX"
24183. .25237 /evidence=not experimental
repeat_region /rpt_family="THE1C-internal"
25244. .25614 /evidence=not experimental
repeat_region /rpt_family="THE1C"
25635. .25687 /evidence=not experimental
repeat_region /rpt_family="AT_rich"
complement(26553..26865) /evidence=not experimental
repeat_region /rpt_family="AluX"
27375. .27563 /evidence=not experimental
repeat_region /rpt_family="MIR"
27567. .27675 /evidence=not experimental
repeat_region /rpt_family="CTA)n"
complement(27988..28357) /evidence=not experimental
repeat_region /rpt_family="THE1C"
28375. .28482 /evidence=not experimental
repeat_region /rpt_family="MER5A"
28724. .28931 /evidence=not experimental
repeat_region /rpt_family="MIR"
28942. .28975 /evidence=not experimental
repeat_region /evidence=not experimental
/rpt_family="CAT)n"
29181. .29201 /evidence=not experimental
repeat_region /rpt_family="AT_rich"
29280. .29520 /evidence=not experimental
repeat_region /rpt_family="MER46A"
complement(29587..29892) /evidence=not experimental
repeat_region /rpt_family="AluY"
29930. .30246 /evidence=not experimental
repeat_region /rpt_family="AluX"
30323. .30362 /evidence=not experimental
repeat_region /rpt_family="AT_rich"
30565. .30588 /evidence=not experimental
repeat_region /rpt_family="AT_rich"
31034. .31101 /evidence=not experimental

```

```

Query Match 83.2%; Score 20.8; DB 8; Length 151991;
Best Local Similarity 91.7%; Pred. No. 7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAAAGCTGATCT 24
Db 94346 AAAAAAAAAAAGCTGATCT 94323

```

```

RESULT 48
AC144536 152224 bp DNA linear PRI 13-MAY-2003
LOCUS Homo sapiens 3 BAC RP11-213K17 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
ACCESSION AC144536
VERSION AC144536.4 GI:30579134
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 152224)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Blimace,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Deigado,O., Dem,A.L., Ding,Y., Dim,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gorrell,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunarane,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homel,F., Howard,S., Huber,J., Huiyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliver,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,B., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mobhat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Ogih,J., Okuwon,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M.,
Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shooshari,N., Sisson,I., Sodergren,E.,
Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabori,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Umani,K., Vasquez,L.,
Vera,V., Villalob,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 152224)
Worley,K.C.
Direct Submission
Submitted (28-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 152224)
Worley,K.C.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 152224)
Worley,K.C.
Direct Submission

```

JOURNAL

COMMENT

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:30466985.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

SOURCE

Location/Qualifiers
1. 152224
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-213K17"
802. 1120
/rpt_family="AluYo"
1298. 1613
/rpt_family="AluXs"
1668. 1740
/rpt_family="MIRb"
1794. 2079
/rpt_family="AluSp"
3301. 3459
/rpt_family="MLT1H"
3736. 3793
/rpt_family="AT-rich"
3850. 3873
/rpt_family="HAL1"
4123. 4216
/rpt_family="L1MC1"
4677. 4967
/rpt_family="AluYd2"
complement(5181..5272)
repeat_region
complement(5273..5569)
repeat_region
complement(5570..5687)
repeat_region

/rpt_family="L1ME3B"
complement(5815..6001)
/rpt_family="L1M3"
complement(6050..6117)
/rpt_family="MER41B"
6169..6482
/rpt_family="AluXs"
complement(6504..6660)
/rpt_family="AluYb"
6661..6715
/rpt_family="Tg)n"
complement(6716..6841)
/rpt_family="AluYb"
complement(7327..7476)
/rpt_family="HAL1"
complement(7572..7808)
/rpt_family="L1MEDd"
complement(7888..8021)
/rpt_family="L1MEDd"
8045..8414
/rpt_family="LTR40c"
8530..8561
/rpt_family="(TTTTG)n"
complement(10299..10380)
/rpt_family="L1MEDd"
11578..11825
/rpt_family="AluYb"
complement(11936..12011)
/rpt_family="LTR33A"
12045..12428
/rpt_family="MSTRB1"
12494..12705
/rpt_family="AluYo"
14092..14314
/rpt_family="MIRb"
14515..15238
/rpt_family="L1MC5"
complement(15261..15497)
/rpt_family="L1MC4a"
15541..16031
/rpt_family="L1MC4"
16036..16150
/rpt_family="MER65A"
complement(16151..16513)
/rpt_family="THER1B"
16514..16804
/rpt_family="MER65A"
16809..17080
/rpt_family="L1MC4"
complement(17331..17477)
/rpt_family="MIR3"
17541..17866
/rpt_family="LTR16A"
18072..18181
/rpt_family="FLAN A"
complement(18185..18501)
/rpt_family="AluXg"

Query Match 83.2%; Score 20.8; DB 8; Length 152224;
Best Local Similarity 91.7%; Pred. No. 7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 AAAAAAAAACTAAAGCTTGATCT 24
81646 AAAAAAAAAAAAAAAAAAGCTTGATCT 81623

RESULT 49
AL831773/c 154298 bp DNA linear ROD 24-SEP-2002
LOCUS AL831773
DEFINITION Mouse DNA sequence from clone RP23-296B1 on chromosome 4, complete
sequence.
ACCESSION AL831773

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1974
 Center clone name: L15_A_18

We are submitting the last 157,0 kilobases of the project at the T7
 end to GenBank at this time. The remainder overlaps AC105015 (MIGR
 project L22977).

FEATURES

Source
 Location/Qualifiers
 1..157030
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-115A18"
 /clone_id="RPC1-11 Human Male BAC"
 complement(2937..3103)
 repeat_region
 /rpt_family="FRAM"
 3637..3656
 /rpt_family="AT_rich"
 complement(3701..3918)
 /rpt_family="AluDb"
 complement(3971..4119)
 /rpt_family="L1M3"
 complement(4806..4994)
 /rpt_family="MER53"
 5613..5752
 /rpt_family="FRAM"
 complement(5913..6218)
 /rpt_family="AluSg"
 complement(6792..7032)
 /rpt_family="MIR"
 8356..8583
 /rpt_family="(T)n"
 complement(9621..9798)
 /rpt_family="MIR"
 10444..10780
 /rpt_family="MER46B"
 complement(112833..13028)
 /rpt_family="L2"
 13938..14007
 /rpt_family="MIR3"
 15186..15268
 /rpt_family="r2"
 complement(15816..16128)
 /rpt_family="AluSx"
 16162..16197
 /rpt_family="AT_rich"
 complement(16198..16483)
 /rpt_family="L1PA16"
 16980..17014
 /rpt_family="AT_rich"
 complement(19311..19616)
 /rpt_family="AluSx"
 20517..20902
 /rpt_family="MLT1B"
 complement(23382..23460)
 /rpt_family="MIR3"
 complement(24064..24221)
 /rpt_family="L1PA7"
 complement(25141..25213)
 /rpt_family="L2"
 25307..25485
 /rpt_family="MIR"
 27373..27439
 /rpt_family="FLAM_A"
 27440..27464
 /rpt_family="(GGA)n"
 27465..27648
 /rpt_family="(GAA)n"

repeat_region 28688..28724
 /rpt_family="(TTTA)n"
 complement(28725..29091)
 /rpt_family="L1MA7"
 repeat_region 29628..29673
 /rpt_family="(T)n"
 31045..31081
 /rpt_family="C-rich"
 31091..31258
 /rpt_family="MIR"
 complement(31619..31870)
 /rpt_family="L1MB2"
 complement(31871..32100)
 /rpt_family="L1PB3"
 complement(32389..32505)
 /rpt_family="L2"
 32910..32959
 /rpt_family="(TA)n"
 33641..33849
 /rpt_family="MER112"
 complement(33992..34311)
 /rpt_family="AluO"
 complement(34315..36131)
 /rpt_family="L1PA16"
 36132..36164
 /rpt_family="U6"
 complement(36165..36366)
 /rpt_family="L1PA16"
 36588..36748
 /rpt_family="MIR"
 complement(36866..37039)
 /rpt_family="MLT1A1"
 complement(37045..37121)
 /rpt_family="MLT1A1"
 37259..40014
 /rpt_family="HERV16"
 38421..38453
 /note="single clone coverage"
 complement(40340..41569)
 /rpt_family="L1PA3"
 41570..42424
 /rpt_family="L1PA3"
 42453..42679
 /rpt_family="HERV16"
 repeat_region
 /rpt_family="HERV16"
 38421..38453
 /note="single clone coverage"
 complement(40340..41569)
 /rpt_family="L1PA3"
 41570..42424
 /rpt_family="L1PA3"
 42453..42679
 /rpt_family="HERV16"

Query Match 83.2%; Score 20.8; DB 8; Length 157030;
 Best Local Similarity 91.7%; Pred. No. 6.9e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 AAAAAAAAACTAAGCTGATCT 24
 Db 121711 AAAAAAAAAAAAGCTGATCT 121688

Search completed: December 14, 2005, 11:10:29
 Job time : 875.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 ; Search time 203.2 Seconds

(without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-3

Perfect score: 25
Sequence: 1 aaaaaaaaaaactgaagcttgatcct 25Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 150 summaries

Database: N_Geneseq_21.*

1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004as:*
13: Geneseq2004bs:*
14: Geneseq2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.4	85.6	1666	13 ADX31754	Adx31754 Plant full
2	20.8	83.2	13606	4 AAS45458	Aas45458 Chemical
3	20.8	83.2	13606	4 AAS45652	Aas45652 Tumour su
4	20.8	83.2	13606	6 ABL33811	Abi33811 Human imm
5	20.8	83.2	13606	6 ABL28314	Abk28314 DNA trans
6	20.2	80.8	424	5 ABY04081	ABY04081 Human pro
7	20.2	80.8	4254	4 ABL04842	Abi04842 Drosophila
8	20.2	80.8	6051	12 ADQ21014	Adq21014 Human sof
9	20.2	80.8	6746	4 ABL03756	Abi03756 Drosophila
10	20.2	80.8	8333	4 AAS45406	Aas45406 Chemical
11	20.2	80.8	8333	6 ABL33502	Abi33502 Human imm
12	20.2	80.8	8333	6 ABL28255	Abk28255 DNA trans
13	20.2	80.8	9483	6 ABL32377	Abi32377 Human imm
14	20.2	80.8	9483	6 ABL70516	Abi70516 Chemical
15	20.2	80.8	9483	6 AAS61092	Aas61092 Human gen
16	20.2	80.8	26230	11 ACN44678	Acn44678 Human gen
17	20.2	80.8	32433	9 ADQ02630	Adq02630 Human FLT
18	20.2	80.8	32433	10 ADB72368	ADB72368 Human FLT
19	20.2	80.8	32433	10 ADE95878	Ade95878 Human FLT

C 20	20.2	80.8	39401	14 AD213218	Ad213218 Murine ca
C 21	20.2	80.8	41991	13 ABD32684	ABD32684 Human can
C 22	20.2	80.8	61020	4 AAS46788	Aas46788 Tumour su
C 23	20.2	80.8	62154	12 ADQ97611	Adq97611 Human can
C 24	20.2	80.8	110000	8 ABX16390	ABX16390
C 25	20.2	80.8	117730	14 AD212550	Ad212550 Human can
C 26	20.2	80.8	235070	11 ACN45174	Acn45174 Human gen
C 27	19.8	79.2	41	12 ADN89686	Adn89686 Human 202
C 28	19.8	79.2	42	3 AAS37946	Aas37946 DNA synth
C 29	19.8	79.2	42	12 ADL19444	Adl19444 125P5C8 g
C 30	19.8	79.2	43	3 AAZ94401	Aaz94401 CDNA synt
C 31	19.8	79.2	43	3 AAZ93041	Aaz93041 Primer us
C 32	19.8	79.2	43	3 AAZ94895	Aaz94895 CDNA synt
C 33	19.8	79.2	43	3 AAI14804	Aai14804 Oligonuc
C 34	19.8	79.2	43	3 AAA09164	Aaa09164 CDNA synt
C 35	19.8	79.2	43	3 AAD06225	Aad06225 Human SGP
C 36	19.8	79.2	43	4 AAD04804	Aad04804 Human sec
C 37	19.8	79.2	43	4 AAF76005	Aaf76005 RT primer
C 38	19.8	79.2	43	6 ABL50400	Abi50400 Human 158
C 39	19.8	79.2	43	6 ABL50412	Abi50412 Human 158
C 40	19.8	79.2	43	6 ABA03602	Abao3602 Human cDN
C 41	19.8	79.2	43	6 AAS99436	Aas99436 DPNCN CD
C 42	19.8	79.2	43	6 ABK67415	Abk67415 Human 83P
C 43	19.8	79.2	43	6 AAL53469	Aal53469 Zinc tran
C 44	19.8	79.2	43	8 ABZ78169	Abz78169 CDNA synt
C 45	19.8	79.2	43	8 ABZ20556	Abz20556 Cancer as
C 46	19.8	79.2	43	10 ADC71176	Adc71176 DPNCN synt
C 47	19.8	79.2	43	10 ADB84526	Adb84526 DPNCN CD
C 48	19.8	79.2	43	10 ABV99869	Abv99869 Human 121
C 49	19.8	79.2	43	11 AD197735	Adi97735 238P182 g
C 50	19.8	79.2	43	12 ADK83217	Adk83217 191P4D12 g
C 51	19.8	79.2	43	12 ADM63820	Adm63820 Human can
C 52	19.8	79.2	43	12 ADN49850	Adn49850 238P182 c
C 53	19.8	79.2	43	12 ADQ64540	Adq64540 Human 213
C 54	19.8	79.2	43	12 ADQ68089	Adq68089 Cancer re
C 55	19.8	79.2	44	5 AAS11665	Aas11665 Prostate
C 56	19.8	79.2	107	3 AAI14809	Aai14809 Fragment
C 57	19.8	79.2	124	8 ABZ20567	Abz20567 Cancer as
C 58	19.8	79.2	192	8 ABZ78104	Abz78104 Human sup
C 59	19.8	79.2	192	13 ADU98925	Adu98925 Human 109
C 60	19.8	79.2	346	8 ABZ20564	Abz20564 Cancer as
C 61	19.8	79.2	425	5 AAS11664	Aas11664 84P2A9 su
C 62	19.8	79.2	945	10 ADF03876	Adf03876 Bacterial
C 63	19.8	79.2	2000	6 ABZ14945	Abz14945 Arabidops
C 64	19.4	77.6	28564	10 ADD47028	Add47028 Human gen
C 65	19.2	76.8	38	2 AAT77235	Aat77235 Rat fibro
C 66	19.2	76.8	38	3 AAZ89341	Aaz89341 Human UCP
C 67	19.2	76.8	49	6 ABZ12149	Abz12149 Colour Fa
C 68	19.2	76.8	50	2 AAV17022	Aav17022 Telomeras
C 69	19.2	76.8	51	8 ABV74551	Abv74551 Murine PR
C 70	19.2	76.8	52	2 AAT90435	Aat90435 Alpha-gal
C 71	19.2	76.8	52	2 AAT72983	Aat72983 Sugar bee
C 72	19.2	76.8	52	2 AAZ30963	Aaz30963 Porcine C
C 73	19.2	76.8	52	2 AAZ28224	Aaz28224 Synthetic
C 74	19.2	76.8	52	3 AAA29293	Aaa29293 QT primer
C 75	19.2	76.8	52	3 AAA62638	Aaa62638 Bacterial
C 76	19.2	76.8	52	3 AAA62637	Aaa62637 Bacterial
C 77	19.2	76.8	52	3 AAA63834	Aaa63834 Anchor pr
C 78	19.2	76.8	52	4 AAC91809	Aac91809 Human HMG
C 79	19.2	76.8	52	5 AAC86119	Aac86119 Primer UN
C 80	19.2	76.8	52	5 AAH88117	Aah88117 Mencha hy
C 81	19.2	76.8	52	5 ABZ25558	Abz25558 Murine in
C 82	19.2	76.8	53	10 ADD90049	Add90049 Primer QT
C 83	19.2	76.8	115	7 ADS72885	Ads72885 Human kid
C 84	19.2	76.8	115	7 ADW41739	Adw41739 CDNA elev
C 85	19.2	76.8	135	5 ABV62010	Abv62010 Human pro
C 86	19.2	76.8	208	4 AAS60053	Aas60053 Human can
C 87	19.2	76.8	226	2 AAT43285	Aat43285 Promoter
C 88	19.2	76.8	334	14 ADY99568	Ady99568 T. reesei
C 89	19.2	76.8	351	14 ADY99260	Ady99260 T. reesei
C 90	19.2	76.8	387	5 ABK44929	Abk44929 CDNA enco
C 91	19.2	76.8	387	5 ABV07913	Abv07913 Human pro
C 92	19.2	76.8	400	5 ABV07451	Abv07451 Human pro

C 93	19.2	76.8	412	10	ABT21658	Abt21658 Breast ca
C 94	19.2	76.8	489	4	AAK59934	AAK59934 Human imm
C 95	19.2	76.8	521	5	ABV57594	ABV57594 Human pro
C 96	19.2	76.8	531	14	ADY98784	Ady98784 T_reesei
C 97	19.2	76.8	650	11	ACN64624	Acn64624 Breast ca
C 98	19.2	76.8	669	3	AAZ80258	Aaz80258 Human col
C 99	19.2	76.8	943	10	ADH02884	Adh02884 Fuchsia F
C 100	19.2	76.8	1999	10	ADCO8455	Adco8455 Rice DNA
C 101	19.2	76.8	2000	7	ADZ75357	Adz75357 Rice prom
C 102	19.2	76.8	2000	12	ADU41650	Adj41650 Plant CDN
C 103	19.2	76.8	3870	14	ADV99935	Adv99935 Human PEM
C 104	19.2	76.8	5096	5	AAS32234	Aas32234 Human DNA
C 105	19.2	76.8	5096	6	ABS67537	Abse67537 Novel hum
C 106	19.2	76.8	6048	6	ABL32509	Ab132509 Human imm
C 107	19.2	76.8	6852	6	ABL70311	Ab170311 Chemical1
C 108	19.2	76.8	6852	6	AAS61249	Aas61249 Human gen
C 109	19.2	76.8	6904	6	ABL32212	Ab132212 Human imm
C 110	19.2	76.8	6977	4	AAS46627	Aas46627 Tumour su
C 111	19.2	76.8	8576	6	ABL34229	Ab134229 Human imm
C 112	19.2	76.8	9139	3	AAZ35274	Aaz35274 Soybean r
C 113	19.2	76.8	13001	12	AD015305	Adq15305 Human thy
C 114	19.2	76.8	13673	4	AAK76548	Aak76548 Human imm
C 115	19.2	76.8	13673	4	AAK81195	Aak81195 Human imm
C 116	19.2	76.8	13966	4	AAS31523	Aas31523 Human DNA
C 117	19.2	76.8	13966	6	ABQ66847	Abq66847 Human pol
C 118	19.2	76.8	13966	10	ADC11134	Adc11134 Human DNA
C 119	19.2	76.8	14001	6	ABQ66846	Abq66846 Human pol
C 120	19.2	76.8	14001	6	ABQ66846	Abq66846 Human pol
C 121	19.2	76.8	14001	10	ADC11133	Adc11133 Human DNA
C 122	19.2	76.8	14712	6	ABN80249	Abn80249 Human che
C 123	19.2	76.8	28860	4	ABL09806	Ab109806 Drosophila
C 124	19.2	76.8	30013	4	AAS41960	Aas41960 Genomic s
C 125	19.2	76.8	30013	4	AAI36932	Aai36932 Human mus
C 126	19.2	76.8	30013	8	ABX59920	Abx59920 CDNA enco
C 127	19.2	76.8	30013	12	ADU30670	Adj30670 Human mus
C 128	19.2	76.8	51321	13	ABD32684	Abd32684 Human can
C 129	19.2	76.8	51321	13	ABD32784	Abd32784 Mouse can
C 130	19.2	76.8	54548	3	AAZ45596	Aaz45596 DNA seque
C 131	19.2	76.8	110000	10	ADG70184	Adg70184 DNA of BA
C 132	19.2	76.8	110000	11	ACN43984	Acn43984 Mouse gen
C 133	19.2	76.8	110000	11	ACN43984	Acn43984 Mouse gen
C 134	19.2	76.8	110000	11	ACN44014	Acn44014 2 of
C 135	19.2	76.8	110000	13	ABD32968	Abd32968 3 of
C 136	19.2	76.8	110000	13	ABD32968	Abd32968 4 of
C 137	19.2	76.8	145831	6	ABL69213	Ab169213 Prostate
C 138	19.2	76.8	145831	6	ABL68806	Ab168806 Lung canc
C 139	19.2	76.8	145831	6	ABL68588	Ab168588 Kidney ca
C 140	19.2	76.8	145831	6	ABL62309	Ab162309 Cotton ade
C 141	19.2	76.8	145831	6	ABT10149	Abt10149 Human bre
C 142	19.2	76.8	154875	14	AAE61197	Aae61197 Human MYH
C 143	19.2	76.8	175590	10	ADD50650	Add50650 BMC seque
C 144	19.2	76.8	175590	14	ADV77908	Adv77908 Human BAC
C 145	19.2	76.8	196686	11	ACN44170	Acn44170 Human gen
C 146	19.2	76.8	217409	11	ACN45150	Acn45150 Human gen
C 147	18.8	75.2	567	12	ACH71128	Ach71128 Human gen
C 148	18.8	75.2	686	6	ABQ29799	Abq29799 Oligonuc1
C 149	18.8	75.2	686	6	ABQ29798	Abq29798 Oligonuc1
C 150	18.8	75.2	902	6	ABQ36622	Abq36622 Oligonuc1

ALIGNMENTS

RESULT 1
ADX31754
ID ADX31754 standard; cDNA; 1666 BP.

AC ADX31754;

DT 21-APR-2005 (first entry)

DE Plant full length insert polynucleotide seqid 14574.

KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.

OS Unidentified.

XX US2004034888-A1.

XX 19-FEB-2004.

XX 28-APR-2003; 2003US-00425114.

XX 06-MAY-1999; 99US-00304517.

XX 05-NOV-2001; 2001US-00985678.

XX (LIU/) LIU J.

XX (ZHOU/) ZHOU Y.

XX (KOVA/) KOVALIC D K.

XX (SCRE/) SCREEN S E.

XX (TABA/) TABASKA J E.

XX (CAO/) CAO Y.

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

XX WPI; 2004-180133/17.

XX The invention describes a recombinant DNA construct comprising a

XX polynucleotide consisting of a sequence encoding an amino acid sequence

XX available in electronic form from the US patent office at

XX ftp://seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide

XX of the invention are also useful in physical arrays of molecules and as

XX plant breeding markers. The recombinant DNA construct is useful for

XX improving plant tolerance to cold, heat, drought, herbicides, extreme

XX osmotic conditions, pathogens or pests, for manipulating growth rate in

XX plant cells by modification of the cell cycle pathway, for conferring

XX lignin or plant growth regulators, for producing galactomannan,

XX recombination in plants, for improving yield by modification of

XX photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake

XX or by providing improved plant growth and development under at least one

XX stress condition or for modifying seed oil or protein yield and/or

XX content. This sequence represents a plant full length insert

XX polynucleotide that can be used in the recombinant DNA construct of the

XX invention.

XX Sequence 1666 BP; 532 A; 259 C; 429 G; 446 T; 0 U; 0 Other;

XX Query Match 85.6%; Score 21.4; DB 13; Length 1666;

XX Best Local Similarity 95.7%; Pred. No. 2e+02; Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAAACTAAGCTTGATCTT 25

Db 1643 AAAAAAGACTTAAGCTTGATCTT 1665

RESULT 2

AA545458/c

AC AA545458;

XX 18-DEC-2001 (first entry)
 XX Chemically pretreated genomic DNA associated with cell cycle #82.
 XX
 XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; Levy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;
 KW immunosuppressive; antitumour; cytosolic; antiarteriosclerotic; ds;
 KW PCR primer.
 XX
 XX Homo sapiens.
 XX
 XX WO200168911-A2.
 XX
 XX 20-SEP-2001.
 XX
 XX 15-MAR-2001; 2001WO-EP002945.
 XX
 XX 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 XX WPI; 2001-602751/68.
 DR
 XX
 XX Designing primers and probes for analyzing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle.
 XX
 XX Claim 1; SEQ ID NO 163; 28bp; English.
 XX
 XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Levy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers
 CC
 XX
 SQ Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
 Query Match 83.2%; Score 20.8; DB 4; Length 13606;
 Best Local Similarity 91.7%; Pred. No. 3.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTAAAGCTGATCT 24
 Db 5439 AAAAAAAAACTAAAGCTGATCT 5416
 RESULT 3
 AAS4562/C
 ID AAS4562 standard; DNA; 13606 BP.
 XX
 XX AAS4562;
 XX
 XX 18-DEC-2001 (first entry)
 XX

DE Tumour suppressor gene derived chemically modified sequence #284.
 XX
 XX Human; tumour suppressor gene; oncogene; antitumour; cytosolic; cancer;
 KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200168912-A2.
 XX
 XX 20-SEP-2001.
 XX
 XX 15-MAR-2001; 2001WO-EP002955.
 XX
 XX 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 XX WPI; 2001-602752/68.
 DR
 XX
 XX Fragments of chemically modified genes associated with tumor suppressor
 PT genes and oncogenes, useful in designing primers and probes for analyzing
 PT diseases associated with cytosine methylation state e.g. cancer.
 XX
 XX Claim 1; SEQ ID NO 284; 27bp; English.
 XX
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and oncogenes
 CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
 CC 500 are missing from the sequence listing) sequences (5s) and sequences
 CC complementary to (5s). The nucleic acid may be a peptide nucleic acid-
 CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
 CC probes for detecting the cytosine methylation state and/or single
 CC nucleotide polymorphisms and also to be used in an array for analysing
 CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
 CC probes can also be used in a method for ascertaining genetic and/or
 CC epigenetic parameters for the diagnosis and/or therapy of existing
 CC diseases or the predisposition to specific diseases, by analysing
 CC cytosine methylations. The parameters may be compared to another set of
 CC genetic and/or epigenetic parameters, the differences serving as basis
 CC for diagnosis and/or prognosis events which are disadvantageous to
 CC patients. The present sequence is one of the 533 genomic sequences
 CC derived from tumour suppressor genes and oncogenes. Sequences with even
 CC numbered Seq ID numbers are the complementary sequence of the
 CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
 CC 535, except for those whose partner sequence is missing). Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
 Query Match 83.2%; Score 20.8; DB 4; Length 13606;
 Best Local Similarity 91.7%; Pred. No. 3.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTAAAGCTGATCT 24
 Db 5439 AAAAAAAAACTAAAGCTGATCT 5416
 RESULT 4
 ABL33811/C
 ID ABL33811 standard; DNA; 13606 BP.
 XX
 XX ABL33811;
 AC

```

XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1784.
DE
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; anti-anemic; cytostatic; noctropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antirheumatic; antidiabetic; antiparasitic;
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1784; 32pp + Sequence listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
SQ
Query Match 83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 5439 AAAAAAAAACTAAAGCTTGATCT 5416

```

```

XX neurodegenerative disorder; Wardenburg syndrome; Niemann-Pick disease;
XX myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
XX angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
XX polyglutamine disorder; solid tumour.
XX
XX Unidentified.
XX
XX WO200192565-A2.
XX
XX 06-DEC-2001.
XX
XX 06-APR-2001; 2001WO-EP003973.
XX
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-090046/12.
XX
XX New nucleic acids or oligomers, useful for diagnosing or treating
XX diseases associated with DNA transcription, e.g. immunological disorders,
XX Werner syndrome, psoriasis, myocardial infarction, solid tumors or
XX cancer.
XX
XX Claim 1; SEQ ID NO 188; 32pp; English.
XX
XX The invention relates to a nucleic acid, which comprises a segment of the
XX chemically pretreated DNA of genes associated with DNA transcription from
XX one of 346 sequences, and an oligomer, in particular an oligonucleotide
XX or peptide nucleic acid (PNA)-oligomer that hybridizes to or is identical
XX to the chemically pretreated DNA of genes associated with DNA
XX transcription. The set of oligomer probes are useful for detecting the
XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
XX in a chemically pretreated genomic DNA. The nucleic acids are useful for
XX diagnosing or treating diseases associated with DNA transcription
XX (particularly with the methylation status), e.g. adenoma desminase
XX deficiency, viral infection, retroviral infection, Sezary syndrome,
XX hematological disorders, immunological disorders, Werner syndrome,
XX tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
XX neurological disorders, neurodegenerative disorders, Wardenburg
XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
XX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumors
XX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
XX associated genomic DNA molecules of the invention. Note: The sequence
XX data for this patent did not form part of the printed specification but
XX was obtained in electronic format directly from the European Patent
XX Office
XX
XX Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
SQ
Query Match 83.2%; Score 20.8; DB 6; Length 13606;
Best Local Similarity 91.7%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 5439 AAAAAAAAACTAAAGCTTGATCT 5416

```

```

RESULT 5
ABK28314/C
ID ABK28314 standard; DNA; 13606 BP.
XX
XX ABK28314;
XX
XX 23-APR-2002 (first entry)
XX
XX DNA transcription associated complementary genomic DNA #94.
XX
XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
XX PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
XX single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
XX viral infection; Sezary syndrome; haematological disorder; tuberculosis;
XX immunological disorder; Werner syndrome; developmental disorder;
XX psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
XX

```

```

RESULT 6
ABV04081/C
ID ABV04081 standard; cDNA; 424 BP.
XX
XX ABV04081;
XX
XX 13-SEP-2002 (first entry)
XX

```


DE Human prostate expression marker CDNA 4072.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX Homo sapiens.
OS
XX WO200160860-A2.
PN
XX
XX 23-AUG-2001.
PD
XX
XX 20-FEB-2001; 2001WO-US005171.
PF
XX
XX 17-FEB-2000; 2000US-0183319P.
PR
XX 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
PR
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
PI Schlegel R, Endege WO, Monahan JE;
PI
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 716; 11750pp; English.
PS
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
CC
XX
SQ Sequence 424 BP; 101 A; 99 C; 88 G; 133 T; 0 U; 3 Other;

Query Match 80.8%; Score 20.2; DB 5; Length 424;
Best Local Similarity 88.0%; Pred. No. 5.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAGCTTGGATCTT 25
Db 374 AAAAAAAAAAAGCTTGGATCTT 350

RESULT 7
ABL04842/c
ID ABL04842 standard; CDNA; 4254 BP.
XX
XX ABL04842;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 9008.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX

PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
PF
XX
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
PR
XX
XX (PERE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX P-PSDB; ABB60739.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX
PS Claim 1; SEQ ID NO 9008; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 4254 BP; 1253 A; 999 C; 1013 G; 989 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 4; Length 4254;
Best Local Similarity 88.0%; Pred. No. 5.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAGCTTGGATCTT 25
Db 827 AAAAAAAAAAAGCTTGGATCTT 803

RESULT 8
ADQ21014/c
ID ADQ21014 standard; DNA; 6051 BP.
XX
XX ADQ21014;
AC
XX
XX 26-AUG-2004 (first entry)
DT
XX
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 3834.
DE
XX
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM ds.
XX
XX Homo sapiens.
OS
XX
XX WO2004048938-A2.
PN
XX
XX 10-JUN-2004.
PD
XX
XX 26-NOV-2003; 2003WO-US038193.
PF
XX
XX 26-NOV-2002; 2002US-0429739P.
PR
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
PA
XX
XX Aziz N, Ginsburg WM, Zlotnik A;
PI
XX WPI; 2004-441208/41.
DR
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample

PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX
 PS Example 2; SEQ ID NO 3834; 210pp; English.
 XX
 CC The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 SQ
 Query Match 80.8%; Score 20.2; DB 12; Length 6051;
 Best Local Similarity 88.0%; Pred. No. 5.6e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAAAGCTGATCTT 25
 Db 4168 AAAAAAAAACTAAAGCTGATCTT 4144
 RESULT 9
 ABL03756/c
 ID ABL03756 standard; cDNA; 6746 BP.
 XX
 AC ABL03756;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5750.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001MO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW,
 XX
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB59653.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 5750; 21pp + Sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6746 BP; 1670 A; 1626 C; 1656 G; 1794 T; 0 U; 0 Other;
 QY
 Query Match 80.8%; Score 20.2; DB 4; Length 6746;
 Best Local Similarity 88.0%; Pred. No. 5.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAAAGCTGATCTT 25
 Db 5953 AAAAAAAAACTAAAGCTGATCTT 5929
 RESULT 10
 AAS45406/c
 ID AAS45406 standard; DNA; 8333 BP.
 XX
 AC AAS45406;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Chemically pretreated genomic DNA associated with cell cycle #56.
 XX
 KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KW PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200168911-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001MO-EP002945.
 XX
 PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPiG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602751/68.
 XX
 DT Designing primers and probes for analyzing diseases associated with
 DT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle.
 XX
 PS Claim 1; SEQ ID NO 111; 28pp; English.
 XX
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,

CC arteriosclerosis, solid tumours and cancers
 XX Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;
 SQ

Query Match 80.8%; Score 20.2; DB 4; Length 8333;
 Best Local Similarity 88.0%; Pred. No. 5.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAGCTTGAATCTT 25
 DB 525 AAAAAAAAAAAGCTTGAATCTT 501

RESULT 11
 ABLJ3502/c
 ID ABLJ3502 standard; DNA; 8333 BP.
 XX
 AC ABLJ3502;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1475.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cyrostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI, 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 1475; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 SQ Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 8333;
 Best Local Similarity 88.0%; Pred. No. 5.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAGCTTGAATCTT 25
 DB 525 AAAAAAAAAAAGCTTGAATCTT 501

RESULT 12
 ABK28255/c
 ID ABK28255 standard; DNA; 8333 BP.
 XX
 AC ABK28255;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE DNA transcription associated genomic DNA #65.
 XX
 KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Wardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.
 XX
 OS Unidentified.
 XX
 PN WO200192565-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 06-APR-2001; 2001WO-EP003973.
 XX
 PR 06-APR-2000; 2000DE-01019058.
 XX
 PR 07-APR-2000; 2000DE-01019173.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI, 2002-090046/12.
 XX
 PT New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological disorders,
 PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
 PT cancer.
 XX
 PS Claim 1; SEQ ID NO 129; 32pp; English.
 XX
 CC The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Wardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification but
 CC was obtained in electronic format directly from the European Patent
 CC Office
 XX
 SQ Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 8333;
 Best Local Similarity 88.0%; Pred. No. 5.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
 |||||
 DB 525 AAAAAAAAACTAAAGCTTGATCTT 501

RESULT 13
 ABL32377/c
 ID ABL32377 standard; DNA; 9483 BP.
 XX
 AC ABL32377;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 350.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipariatic;
 KW antineumatic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW de.

XX Homo sapiens.
 XX WO200200928-A2.
 XX
 PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX

PA (EPIC-) EPIDENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 XX

PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX Claim 1; SEQ ID NO 350; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC

SQ Sequence 9483 BP; 2966 A; 95 C; 1789 G; 4633 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 9483;
 Best Local Similarity 88.0%; Pred. No. 5.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
 |||||
 DB 4182 AAAAAAAAACTAAAGCTTGATCTT 4158

RESULT 14

ABL70516/c
 ID ABL70516 standard; DNA; 9483 BP.
 XX
 AC ABL70516;
 XX

DT 01-JUL-2002 (first entry)
 XX

DE Chemically treated cell signalling DNA sequence complementary to#203.

XX Cell signalling; cytosine methylation; cell signalling disease; cancer;
 KW tumour; cytosolic; ds.
 XX

OS Unidentified.

XX WO200202807-A2.
 XX

PD 10-JAN-2002.
 XX

PF 29-JUN-2001; 2001WO-EP007471.
 XX

PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.

PA (EPIC-) EPIDENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-154758/20.

XX Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signaling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signaling.

XX Claim 1; SEQ ID NO 406; 24pp + Sequence Listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signaling. The activity of the modified sequences of the
 CC invention may be described as cytosolic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signaling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is

CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signaling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signaling. Note: The
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office
 CC

SQ Sequence 9483 BP; 2966 A; 95 C; 1789 G; 4633 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 9483;
 Best Local Similarity 88.0%; Pred. No. 5.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
 |||||
 DB 4182 AAAAAAAAACTAAAGCTTGATCTT 4158

RESULT 15
 AAS61092/c
 ID AAS61092 standard; DNA; 9483 BP.

AC AAS61092;
 XX

DT 29-JAN-2002 (first entry)
 XX

DE Human gene regulation-associated gene oligonucleotide #47.

Human; Gene regulation-associated gene; severe combined immunodeficiency;
cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
renal disease; Preeclampsia; cardiac allograft vascular disease;
colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
immunostimulant; cardioid; antiinflammatory; coagulant; antithrombotic;
nephrotoxic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003968.
XX
PR 06-APR-2000; 2000DE-01019058.
XX
PR 07-APR-2000; 2000DE-01019173.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIDENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K.
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated with
XX gene regulation, useful for analyzing cytosine methylations for diagnosis
XX and therapy of diseases e.g. severe combined immunodeficiency disease.
PS Claim 1; SEQ ID NO 48; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 9483 BP; 2966 A; 95 C; 1789 G; 4633 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 6; Length 9483;
Best Local Similarity 88.0%; Pred. No. 5.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 4182 AAAAAAAAAATAAATCTTATCTT 4158
RESULT 16
ID ACN44678 standard; DNA; 26230 BP.
XX
AC ACN44678;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG1737521.

Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
KM Homo sapiens.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
DR 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 1246; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182986A1, for which no sequence data was published
XX
SQ Sequence 26230 BP; 7368 A; 4959 C; 5216 G; 8570 T; 0 U; 117 Other;
Query Match 80.8%; Score 20.2; DB 11; Length 26230;
Best Local Similarity 88.0%; Pred. No. 5.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 17551 AAAAAAAAAAGAAATCTTGATCTT 17527
RESULT 17
ID ADA02630 standard; DNA; 32433 BP.
XX
AC ADA02630;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human FLT3 carcinoma associated gene, SEQ ID NO:1148.
XX
KM Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.

```
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW;
PI
XX WPI; 2003-587068/55.
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1148; 245bp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 32433 BP; 8463 A; 7444 C; 7079 G; 9447 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 9; Length 32433;
Best Local Similarity 88.0%; Pred. No. 5.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAGCTTAAGCTTATCTT 25
Db 1433 AAAAAAAAAAAGCTTATATT 1409
RESULT 18
ADB72368/C
ID ADB72368 standard; DNA; 32433 BP.
XX
XX ADB72368;
XX
XX 04-DEC-2003 (first entry)
XX
XX Human FLT3 gene.
XX
XX human; de; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KM cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX Homo sapiens.
XX
XX WO2003008583-A2.
XX
XX 30-JAN-2003.
XX
XX 26-DEC-2001; 2001WO-US051291.
XX
XX 02-MAR-2001; 2001US-00798586.
XX
XX 23-OCT-2001; 2001US-00004113.
XX
XX 08-NOV-2001; 2001US-00052482.
XX
XX 30-NOV-2001; 2001US-00997222.
XX
XX 20-DEC-2001; 2001US-00034650.
XX
```

```
PA (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW, Engelhard EK;
PI
XX WPI; 2003-239337/23.
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcoma.
XX
XX Claim 1; SEQ ID NO 196; 2304bp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 32433 BP; 8463 A; 7444 C; 7079 G; 9447 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 10; Length 32433;
Best Local Similarity 88.0%; Pred. No. 5.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAGCTTAAGCTTATCTT 25
Db 1433 AAAAAAAAAAAGCTTATATT 1409
RESULT 19
ADE95878/C
ID ADE95878 standard; DNA; 32433 BP.
XX
XX ADE95878;
XX
XX 12-FEB-2004 (first entry)
XX
XX Human FLT3 gene genomic DNA sequence.
XX
XX cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
KM lymphoma; breast cancer; prostate cancer; leukaemia; de; human; FLT3.
XX
XX Homo sapiens.
XX
XX WO2003039484-A2.
XX
XX 15-MAY-2003.
XX
XX 08-NOV-2002; 2002WO-US036071.
XX
XX 08-NOV-2001; 2001US-00052482.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW, Engelhard EK;
PI
XX WPI; 2003-441462/41.
XX
XX New carcinoma associated nucleic acids and proteins, useful for screening
PT drug candidates, or for diagnosing and treating carcinomas, e.g.
XX lymphoma, breast cancer, prostate cancer or leukemia.
XX
XX Claim 1; SEQ ID NO 136; 793bp; English.
XX
XX This invention relates to novel recombinant nucleic acids for use in
CC diagnosis and treatment of cancer, especially carcinomas, as well as the
CC use of compositions in screening methods. The compositions of the
CC invention may have cytostatic activity whilst the disclosed sequences may
CC be useful for gene therapy. The carcinoma associated nucleic acids and
CC proteins are useful for diagnosing and treating carcinomas, for example
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC drug candidates or bioactive agents capable of binding to, or modulating
```

CC the activity of, a carcinoma associated protein. The present sequence is
 CC the genomic DNA sequence of the human FLT3 gene which is a carcinoma
 CC associated gene of the invention.
 XX
 SQ Sequence 32433 BP; 8463 A; 7444 C; 7079 G; 9447 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 10; Length 32433;
 Best Local Similarity 88.0%; Pred. No. 5.8e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTAAAGCTTGATCTT 25
 Db 1433 AAAAAAAAAAAAGCTTGATATT 1409

RESULT 20
 AD213218/c
 ID AD213218 standard; DNA; 39401 BP.
 XX
 AC AD213218;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE Murine cancer-associated genomic DNA #61.
 XX
 KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
 KW cyrostatic; gene; ds.
 XX
 OS Mus sp.
 XX
 PN WO2005031001-A2.
 XX
 PD 07-APR-2005.
 XX
 PF 23-SEP-2004; 2004WO-US031617.
 XX
 PR 23-SEP-2003; 2003US-00669920.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Morris DW, Malandro MS;
 XX
 DR WPI; 2005-273395/28.
 XX
 PT Nucleic acid array useful for detecting cancer associated nucleic acid,
 PT comprises two or more nucleic acid probes.
 XX
 PS Disclosure: SEQ ID NO 738; 198pp; English.
 XX

The invention relates to a nucleic acid array for detecting a cancer associated (CA) nucleic acid, comprising two or more nucleic acid probes. The invention also relates to a peptide array comprising two or more isolated polypeptides encoded by a CA nucleic acid sequence, a compound that binds to a polypeptide, an isolated antibody or its fragment which binds to a polypeptide, which is prepared by immunizing a host animal with a composition comprising the polypeptide or its antigen binding fragment and collection cells from the host expressing antibodies against the antigen or its antigen binding fragment, a composition comprising the antibody and a carrier, a method of screening for anticancer activity, a method of detecting a CA nucleic acid, a method of diagnosing cancer, a method of treating cancer and a method of inhibiting expression of a CA nucleic acid in a cell. The CA nucleic acids are useful for detecting CA nucleic acids. The antibody is useful for detecting the presence or absence of cancer cells in an individual which involves contacting cells from the individual with the antibody and detecting a complex of a CA protein from the cancer cells and the antibody, where the detection of the complex correlates with the presence of cancer cells in the individual. The composition is useful for inhibiting growth of cancer cells in an individual or for delivering a therapeutic agent to cancer cells in an individual. The invention is also useful for diagnosing cancer, for treating cancer and for inhibiting expression of a CA gene in a cell. This sequence represents murine cancer-associated genomic DNA of the invention.

XX
 SQ Sequence 39401 BP; 10528 A; 8379 C; 9062 G; 10891 T; 0 U; 541 Other;
 Query Match 80.8%; Score 20.2; DB 14; Length 39401;
 Best Local Similarity 88.0%; Pred. No. 5.8e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTAAAGCTTGATCTT 25
 Db 34884 AAAAAAAAACTAAGCTTAGATTTT 34860

RESULT 21
 ABD32684
 ID ABD32684 standard; DNA; 41991 BP.
 XX
 AC ABD32684;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human cancer-associated genomic DNA HD13-117.
 XX
 KW Human; ds; cancer-associated protein; gene; cyrostatic; cancer;
 KW leukemia; lymphoma; CAP.
 XX
 OS Homo sapiens.
 XX
 PN WO2004074320-A2.
 XX
 PD 02-SEP-2004.
 XX
 PF 17-FEB-2004; 2004WO-US004730.
 XX
 PR 14-FEB-2003; 2003US-00367094.
 XX
 PR 14-MAR-2003; 2003US-00388838.
 XX
 PR 15-APR-2003; 2003US-00417375.
 XX
 PR 13-JUN-2003; 2003US-00461862.
 XX
 PR 15-SEP-2003; 2003US-00663431.
 XX
 PR 15-DEC-2003; 2003US-00737318.
 XX
 PA (SAGR-) SAGRES DISCOVERY INC.
 XX
 PI Morris DW, Malandro MS;
 XX
 DR WPI; 2004-652914/63.
 XX
 PT New isolated cancer-associated polynucleotides and polypeptides useful
 PT for diagnosing, preventing or treating cancer, especially lymphoma and
 PT leukemia, or in screening for agents that modulate cancer.
 XX
 PS claim 16; seqid 239; 310pp; English.
 XX

The invention relates to an isolated nucleic acid comprising at least 10 contiguous nucleotides of any of the 233 polynucleotide sequences given in the specification, or its complement. The nucleic acids encode cancer-associated proteins. Also included are an expression vector comprising the isolated nucleic acid cited above, a host cell comprising the above recombinant nucleic acid or expression vector, a microarray for detecting a cancer-associated (CA) nucleic acid comprising at least one probe comprising at least 10 contiguous nucleotides of any of the above-mentioned nucleotide sequences, an isolated polypeptide (encoded within an open reading frame of a CA sequence selected from any of the 95 polynucleotide sequences as mentioned in the specification, or its complement), an isolated antibody, (or its antigen binding fragment) that binds to the above polypeptide, a hybridoma that produces the above monoclonal antibody, a pharmaceutical composition comprising the above antibody and a pharmaceutical excipient, a kit for detecting cancer cells (comprising the antibody cited above, methods for diagnosing cancer or for detecting the presence or absence of cancer cells in an individual, a method for inhibiting growth of cancer cells in an individual, a method for delivering a therapeutic agent to cancer cells in an individual, an electronic library comprising the above polynucleotide or polypeptide (or their fragments), methods of screening

CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 41991 BP; 12474 A; 8216 C; 8278 G; 13023 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 13; Length 41991;
XX Best Local Similarity 88.0%; Pred. No. 5.8e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 35367 AAAAAAAAAATTAAAGCATCATCTT 35391

RESULT 22
AAS46788/c
ID AAS46788 standard; DNA; 61020 BP.
XX
XX AAS46788;
XX
XX 18-DEC-2001 (first entry).
XX
XX Tumour suppressor gene derived chemically modified sequence #514.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytosolic; cancer;
XX tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP002955.
XX
XX 15-MAR-2000; 2000DE-01013847.
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumor suppressor
XX genes and oncogenes, useful in designing primers and probes for analyzing
XX diseases associated with cytosine methylation state e.g. cancer.
XX
XX Claim 1; SEQ ID NO 514; 27bp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The

CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 4; Length 61020;
XX Best Local Similarity 88.0%; Pred. No. 5.8e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 25659 AAAAAAAAAATTATTCCTT 25635

RESULT 23
AD097611
ID AD097611 standard; DNA; 62154 BP.
XX
XX AD097611;
XX
XX 07-OCT-2004 (first entry)
XX
XX Human cancer associated sequence HD10-014, SEQ ID 588.
XX
XX Human Cancer associated sequence HD10-014, SEQ ID 588.
XX
XX Cytosolic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
XX Homo sapiens.
XX
XX WO2004060304-A2.
XX
XX 22-JUL-2004.
XX
XX 22-DEC-2003; 2003WO-US041389.
XX
XX 27-DEC-2002; 2002US-00330773.
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX WPI; 2004-543781/52.
XX
XX New isolated cancer associated nucleic acids comprising at least 10
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX cancers such as leukemia and lymphoma.
XX
XX Claim 1; SEQ ID NO 588; 199bp; English.

CC The present invention relates to cancer associated sequences (AD097025-
CC AD096004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 62154 BP; 18095 A; 14107 C; 13141 G; 16811 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 12; Length 62154;
XX Best Local Similarity 88.0%; Pred. No. 5.9e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;


```
Oy      1 AAAAAAAAACTAAGCTTGATCTT 25
Db      7983 AAAAAAAAAAAAAAAAAAGCTTGATT 8007

RESULT 24
ABX16390.1
Continuation (2 of 7) of ABX16390 from base 100001 (Mouse high growth region.)
WP Sequence split into 7 fragments LOCUS ABX16390 Accession Abx16390
WP Fragment Name      Begin      End
WP ABX16390_0          1      110000
WP ABX16390_1          100001   210000
WP ABX16390_2          200001   310000
WP ABX16390_3          300001   410000
WP ABX16390_4          400001   510000
WP ABX16390_5          500001   610000
WP ABX16390_6          600001   659158

Query Match      80.8%; Score 20.2; DB 8; Length 110000;
Best Local Similarity 88.0%; Pred. No. 5.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy      1 AAAAAAAAACTAAGCTTGATCTT 25
Db      56542 AAAAAAAAAAGACTTAAGCTTGCTGT 56566

RESULT 25
AD212550/C
ID AD212550 standard; DNA; 117730 BP.
XX
XX AD212550;
XX
XX 16-JUN-2005 (first entry)
XX
XX Human cancer-associated genomic DNA #8.
XX
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
XX cytostatic; gene; ds.
XX
XX Homo sapiens.
XX
XX WO2005031001-A2.
XX
XX 07-APR-2005.
XX
XX 23-SEP-2004; 2004WO-US031617.
XX
XX 23-SEP-2003; 2003US-00669920.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Morris DW, Malandro MS;
XX
XX WPI, 2005-273395/28.
XX
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
XX comprises two or more nucleic acid probes.
XX
XX Disclosure; SEQ ID NO 70; 198bp; English.
XX
XX The invention relates to a nucleic acid array for detecting a cancer
XX associated (CA) nucleic acid, comprising two or more nucleic acid probes.
XX The invention also relates to a peptide array comprising two or more
XX isolated polypeptides encoded by a CA nucleic acid sequence, a compound
XX that binds to a polypeptide, an isolated antibody or its fragment which
XX binds to a polypeptide, which is prepared by immunizing a host animal
XX with a composition comprising the polypeptide or its antigen binding
XX fragment and collecting cells from the host expressing antibodies against
XX the antigen or its antigen binding fragment, a composition comprising the
XX antibody and a carrier, a method of screening for anticancer activity, a
XX method of detecting a CA nucleic acid, a method of diagnosing cancer, a
XX method of treating cancer and a method of inhibiting expression of a CA

CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC nucleic acids. The antibody is useful for detecting the presence or
CC absence of cancer cells in an individual which involves contacting cells
CC from the individual with the antibody and detecting a complex of a CA
CC protein from the cancer cells and the antibody, where the detection of
CC the complex correlates with the presence of cancer cells in the
CC individual. The composition is useful for inhibiting growth of cancer
CC cells in an individual or for delivering a therapeutic agent to cancer
CC cells in an individual. The invention is also useful for diagnosing
CC cancer, for treating cancer and for inhibiting expression of a CA gene in
CC a cell. This sequence represents human cancer-associated genomic DNA of
CC the invention.

SQ      Sequence 117730 BP; 32629 A; 25620 C; 25334 G; 34147 T; 0 U; 0 Other;

Query Match      80.8%; Score 20.2; DB 14; Length 117730;
Best Local Similarity 88.0%; Pred. No. 5.9e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy      1 AAAAAAAAACTAAGCTTGATCTT 25
Db      65144 AAAAAAAAAAAAAAAAAAGCTTGATT 65120

RESULT 26
ACN45174/C
ID ACN45174 standard; DNA; 235070 BP.
XX
XX ACN45174;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human genomic sequence hCG15927.
XX
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
XX Homo sapiens.
XX
XX WO2003073826-A2.
XX
XX 12-SEP-2003.
XX
XX 28-FEB-2003; 2003WO-US006235.
XX
XX 01-MAR-2002; 2002US-00087192.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
XX
XX WPI, 2003-328604/31.
XX
XX Claim 1; SEQ ID NO 1990; 0pp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US200218286A1, for which no sequence data was published
XX
XX Sequence 235070 BP; 57319 A; 52466 C; 56014 G; 69049 T; 0 U; 222 Other;
```

Query Match 80.8%; Score 20.2; DB 11; Length 235070;
 Best Local Similarity 88.0%; Pred. No. 5.9e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
 |||||
 DB 126964 AAAAAAAAAAGATTGATCTT 126940

RESULT 27
 ADN89686/c
 ID ADN89686 standard; DNA; 41 BP.

XX AC ADN89686;

XX DT 01-JUL-2004 (first entry)

XX DE Human 202P5A5 cDNA synthesis primer.

XX KW 202P5A5; human; cancer; tumour; ss; primer.

XX OS Homo sapiens.

XX PN WO2004016736-A2.

XX PD 26-FEB-2004.

XX PF 16-JUN-2003; 2003WO-US018996.

XX PR 16-AUG-2002; 2002US-0404306P.

XX PR 01-NOV-2002; 2002US-0423299P.

XX PA (AGEN-) AGENSYS INC.

XX PI Raitano AB, Faris M, Challita-Bid PM, Jakobovits A, Ge W;

XX DR WPI; 2004-203774/19.

XX PT New compositions having the 202P5A5 gene and encoded protein, useful for
 PT diagnosing, preventing, prognosticating or treating cancer of the
 PT prostate, bladder, colon, lung, ovary, breast, stomach, cervix, lymphoma,
 PT bone and/or skin.

XX PS Example 1; SEQ ID NO 28; 266pp; English.

XX CC The invention relates to a composition comprising 202P5A5 proteins. The
 CC composition and proteins are useful for detecting and treating cancer by
 CC inhibiting the growth or viability of cancer cells. The present sequence
 CC represents the human 202P5A5 cDNA synthesis primer.

XX SQ Sequence 41 BP; 3 A; 2 C; 2 G; 34 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 12; Length 41;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATC 23
 |||||
 DB 27 AAAAAAAAAAGATTGATCTT 5

RESULT 28
 AAA37946/c
 ID AAA37946 standard; DNA; 42 BP.

XX AC AAA37946;

XX DT 18-AUG-2000 (first entry)

XX DE DNA synthesis primer used in PTAN gene isolation.

XX KW PTAN; testis specific; prostate cancer; overexpress; chromosome 1q22;

KW diagnose; cancer; breast; vaccine; primer; ss.

XX OS Synthetic.

XX PN WO200020589-A2.

XX PD 13-APR-2000.

XX PF 30-SEP-1999; 99WO-US022985.

XX PR 30-SEP-1998; 98US-0102556P.

XX PR 02-OCT-1998; 98US-0102910P.

XX PR 21-DEC-1998; 98US-0113229P.

XX PR 14-APR-1999; 99US-0129518P.

XX PA (UROG-) UROGENESYS INC.

XX PA (AFAR/) AFAR D E.

XX PA (HUBE/) HUBERT R S.

XX PA (RAIT/) RAITANO A B.

XX PA (MITC/) MITCHELL S C.

XX PI Afar DE, Hubert RS, Raitano AB, Mitchell SC;

XX DR WPI; 2000-317715/27.

XX PT PTAN proteins, and sequences encoding them, used for diagnosing and
 PT treating cancers, especially breast and prostate cancers.

XX PS Example 1; Page 31; 71pp; English.

XX CC This sequence represents a primer used in the isolation of cDNA fragments
 CC of the PTAN (testis specific protein expressed in prostate cancer) gene.
 CC PTAN is expressed in 3 isoforms PTAN-1, 2, and 3. The PTAN gene is
 CC located on chromosome 1q22. PTAN is overexpressed in prostate cancer, and
 CC has a testis specific expression pattern in adult tissues. PTAN shows no
 CC homology to any known gene. PTAN can be used in methods for the diagnosis
 CC of cancer, especially prostate or breast cancer, where the normal tissue
 CC samples are prostate tissue, or breast tissue, bone tissue, lymphatic
 CC tissue, serum, blood, or urine. A vector containing the PTAN nucleotide
 CC sequence, a vaccine composition targeting PTAN, PTAN, ribozymes specific
 CC for PTAN mRNA and antisense sequences, can be used to treat cancer,
 CC especially breast and prostate cancers. Cancer development can be
 CC inhibited by a vaccine composition targeting PTAN

XX SQ Sequence 42 BP; 3 A; 2 C; 2 G; 35 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 3; Length 42;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATC 23
 |||||
 DB 26 AAAAAAAAAAGATTGATC 4

RESULT 29
 ADL19444/c
 ID ADL19444 standard; DNA; 42 BP.

XX AC ADL19444;

XX DT 20-MAY-2004 (first entry)

XX DE 125P5C8 gene-related cDNA synthesis primer.

XX KW 125P5C8; cancer-associated gene; cancer-associated protein; cancer;
 KW cDNA synthesis; primer; PCR; ss.

XX OS Unidentified.

XX PN US2003219444-A1.

XX PD 27-NOV-2003.

XX 13-MAR-2002; 2002US-00099460.
 PF 14-MAR-2001; 2001US-00809638.
 XX
 XX (FARI/) PARIS M.
 PA (CHAL/) CHALITA-EID P M.
 PA (HUBE/) HUBERT R S.
 PA (AFAR/) AFAR D E H.
 PA (RAIT/) RAITANO A B.
 PA (GEWM/) GE W.
 PA (MORR/) MORRISON R K.
 PA (MORR/) MORRISON K J M.
 PA (JAKO/) JAKOBOVITS A.
 PI Faris M, Chalita-Eid PM, Hubert RS, Afar DEH, Raitano AB, Ge W;
 PI Morrison RK, Morrison KJM, Jakobovits A;
 XX WPI; 2004-021932/02.
 DR
 XX New composition comprising a substance that modulates the status of
 PT 125P5C8 gene or a molecule that is modulated by 125P5C8, useful for
 PT diagnosing or treating cancer.
 XX
 PS Example 1; SEQ ID NO 714; 183pp; English.
 XX
 CC The invention comprises a composition which contains a substance that can
 CC modulate the status of 125P5C8 (125P5C8 is a novel cancer-associated
 CC gene/protein), or a molecule that is modulated by 125P5C8. The
 CC composition of the invention is useful for diagnosing or treating cancer.
 CC The present DNA sequence represents a cDNA synthesis primer that was used
 CC in an example of the invention to isolate a fragment of the 125P5C8 gene.
 XX
 SQ Sequence 42 BP; 3 A; 2 C; 2 G; 35 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 12; Length 42;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAAGCTTGATC 23
 DB 26 AAAAAAAAAAAGCTTGATC 4

RESULT 30
 ID AA294401/C
 XX AA294401 standard; DNA; 43 BP.
 AC AA294401;
 XX
 DT 03-JUL-2000 (first entry)
 XX
 DE cDNA synthesis primer DPNCDN.
 XX
 KM PHELIx; human; testis-specific; transcription factor; prostate cancer;
 KM bladder cancer; ovary cancer; testicular cancer; gene therapy; diagnosis;
 KM vaccine; primer; ss.
 XX
 OS Synthetic.
 XX
 PN WO200012709-A2.
 XX
 PD 09-MAR-2000.
 XX
 PF 31-AUG-1999; 99WO-US020137.
 XX
 PR 31-AUG-1998; 98US-0098610P.
 PR 31-OCT-1998; 98US-0106524P.
 XX
 PA (UROG-) UROGENESYS INC.
 PA (AFAR/) AFAR D E.
 PA (HUBE/) HUBERT R S.
 PA (RAIT/) RAITANO A B.

XX Afar DE, Hubert RS, Raitano AB;
 PI WPI; 2000-237872/20.
 XX
 XX Testis specific Helix Loop Helix proteins expressed in cancers and useful
 PT for the prevention, diagnosis and treatment of prostate, bladder and
 PT ovarian tumors.
 XX
 PS Example 1; Page 31; 62pp; English.
 XX
 CC The present sequence is that of cDNA synthesis primer DPNCDN, which was
 CC used in a suppression subtractive hybridization protocol designed to
 CC identify novel prostate and prostate cancer specific genes. cDNAs from
 CC androgen dependent LAPC-4 xenograft were compared with cDNAs derived from
 CC androgen independent LAPC-4 xenograft. A 437 bp clone was obtained. Full-
 CC length cDNA (see AA294275) was subsequently cloned from a testis cDNA
 CC library. This encoded PHELIx (see AA79269), a novel transcription factor
 CC that is normally expressed only in testis tissue, but is up-regulated in
 CC prostate and other types of cancer. The invention provides diagnostic and
 CC therapeutic methods useful in the management of various cancers whic
 CC express PHELIx, including prostate cancer, bladder cancer, ovarian cancer
 CC and testicular cancer
 XX
 SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 3; Length 43;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAAGCTTGATC 23
 DB 27 AAAAAAAAAAAGCTTGATC 5

RESULT 31
 ID AA293041/C
 XX AA293041 standard; DNA; 43 BP.
 AC AA293041;
 XX
 DT 24-JUL-2000 (first entry)
 XX
 DE Primer used for generating human brain specific protein BPC-1 cDNA.
 XX
 KM BPC-1; oncogene; oncogenic; cancer; prostate; bladder; antibody;
 KM antisense; vaccine; detection; prognosis; drug screening; primer; ss.
 XX
 OS Synthetic.
 XX
 PN WO200009691-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 10-AUG-1999; 99WO-US018250.
 PF 10-AUG-1998; 98US-0095982P.
 PR
 XX
 PA (UROG-) UROGENESYS INC.
 PA (AFAR/) AFAR D E.
 PA (HUBE/) HUBERT R S.
 PA (LEON/) LEONG K.
 PA (RAIT/) RAITANO A B.
 PA (SAFF/) SAFFRAN D C.
 PA (JAKO/) JAKOBOVITS A.
 XX
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Safran DC;
 PI Jakobovits A;
 XX
 DR WPI; 2000-206006/18.
 PT
 PT New isolated BPC-1 polypeptides, useful for developing products for the
 diagnosis, staging, prognosis and treatment of cancers, particularly

```

PT prostate or bladder cancer.
XX
XX Example 1; Page 35; 79pp; English.
XX
CC BPC-1 polypeptides and polynucleotides can be used for the detection of
CC BPC-1 polypeptides and polynucleotides in biological samples, this is
CC particularly useful for detecting cancers expressing BPC-1, e.g. prostate
CC cancer or bladder cancer. Antibodies directed against BPC-1 or antisense
CC polynucleotides can be used for treating such cancers. The BPC-1
CC polypeptides can also be used in vaccines for treating or inhibiting the
CC development of a cancer expressing BPC-1. The polypeptides and
CC polynucleotides can also be used for detection, prognosis, drug screening
CC and predicting susceptibility to developing cancer. The BPC-1 polypeptide
CC comprises a CUB domain which is expressed in prostate and bladder
CC carcinoma cells and which shows sequence similarity with CUB domains from
CC other known proteins. In normal human tissues BPC-1 is only expressed in
CC certain tissues of the brain, however, it is expressed at high levels in
CC prostate cancer cells and bladder cancer cells. A number of synthetic
CC oligonucleotides were used to generate BPC-1 cDNA from total cell RNA of
CC tumour cells lines. These primers were a cDNA synthesis primer
CC (AAZ93041), two adaptor sequences (AAZ93042-Z93045), a PCR primer
CC (AAZ93046) and two nested primers (AAZ93047, AAZ93048). This sequence is
CC the cDNA synthesis primer
XX
SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 3; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAACTAAAGCTTGATC 23
Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5
XX
RESULT 32
AAZ94895/c
ID AAZ94895 standard; DNA; 43 BP.
XX
AC AAZ94895;
XX
DT 01-AUG-2000 (first entry)
XX
DE cDNA synthesis primer DPNCN.
XX
KM 22P4F11; human; testis; prostate cancer; diagnosis; gene therapy; marker;
KM vaccine; primer; ss.
XX
OS Synthetic.
XX
PN WO200018925-A1.
XX
PD 06-APR-2000.
XX
PF 30-SEP-1999; 99WO-US023005.
XX
PR 30-SEP-1998; 98US-0102572P.
PR 28-JUL-1999; 99US-0146584P.
XX
PA (UROG-) UROGENESYS INC.
PA (AFAR/) AFAR D E.
PA (HUBE/) HUBERT R S.
PA (MITC/) MITCHELL S C.
XX
PI Afar DE, Hubert RS, Mitchell SC,
XX
DR WPI; 2000-303452/26.
XX
PT Novel testes-specific gene 22P4F11 which is expressed in human prostate
PT cancer and is useful as a diagnostic marker and/or therapeutic target for
PT prostate cancer.
XX
XX Example 1; Page 28; 54pp; English.

```

```

XX
CC The present sequence is that of cDNA synthesis primer DPNCN, which was
CC used in a suppression subtractive hybridisation protocol to identify
CC cDNAs corresponding to genes which may be differentially expressed in
CC human prostate cancer. A partial clone, termed 22P4F11 (see AAZ94894),
CC was obtained and used to identify full-length 22P4F11 cDNA (see
CC AAZ94993). 22P4F11 is a testis-specific gene in normal tissues, and is
CC also expressed in human prostate tumours, in some cases at high levels.
CC The 22P4F11 transcript and/or protein (see AA179489) may represent a
CC useful diagnostic marker and/or therapeutic target for prostate cancer.
CC Methods of using 22P4F11 polynucleotides, polypeptides and antibodies for
CC the diagnosis and treatment of cancers expressing 22P4F11, especially
CC prostate cancer, are provided, as well as vaccines that prevent
CC development of such cancers
XX
SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 3; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAACTAAAGCTTGATC 23
Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5
XX
RESULT 33
AA14804/c
ID AA14804 standard; DNA; 43 BP.
XX
AC AA14804;
XX
DT 08-AUG-2000 (first entry)
XX
DE Oligonucleotide DPNCN used for cDNA synthesis.
XX
KM Prostate cancer; testis-specific protein Y-encoded mRNA; TSPY mRNA;
KM vaccine; primer; ss.
XX
OS Mus sp.
XX
PN WO200020638-A2.
XX
PD 13-APR-2000.
XX
PF 02-OCT-1999; 99WO-US022575.
XX
PR 02-OCT-1998; 98US-0102893P.
XX
PA (UROG-) UROGENESYS INC.
PA (AFAR/) AFAR D E.
XX
PI Afar DE, Hubert RS;
XX
DR WPI; 2000-303803/26.
XX
PT Diagnosing prostate cancer by determining the level of testis-specific
PT protein Y-encoded (TSPY) mRNA or protein and comparing these TSPY mRNA or
PT protein levels to those of a normal tissue sample.
XX
XX Example 1; Page 20; 32pp; English.
XX
CC The specification describes a new method of diagnosis of prostate cancer.
CC The method comprises determining the level of testis-specific protein Y-
CC encoded (TSPY) mRNA or protein, and comparing these TSPY mRNA or protein
CC levels to those of a normal tissue sample. The presence of elevated TSPY
CC mRNA or protein is indicative of prostate cancer. Detection of TSPY mRNA
CC expression or protein levels is useful in the diagnosis of prostate
CC cancer. Antisense polynucleotides complementary to the coding sequence of
CC human TSPY are useful for treating prostate cancer by inhibiting TSPY
CC transcription (when contacted with the TSPY gene) or translation (when
CC contacted with the TSPY mRNA). Ribozymes are also useful for treating
CC prostate cancer by cleaving the TSPY mRNA and therefore inhibiting its

```

CC translation. The vaccine is useful for inhibiting the development of
CC prostate cancer in a patient. The present sequence represents a primer
CC used for CDNA synthesis, used in the course of the invention
XX

SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 3; Length 43;

Best Local Similarity 91.3%; Pred. No. 7.3e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAACTAAGCTTGATC 23
27 AAAAAAAAAAAAAAAAAAGCTTGATC 5

RESULT 34

AAA09164/C

10-AUG-2000 (first entry)

AAA09164;

10-AUG-2000 (first entry)

CDNA synthesis primer for SSH-generation of 36P1A6 gene.

36P1A6; transcription factor; murine EHF homologue; ETS family;

cytostatic; cancer; vaccine; tumorigenesis; ss.

Homo sapiens.

WO200020584-A2.

13-APR-2000.

02-OCT-1999; 99WO-US022576.

02-OCT-1998; 98US-0102744P.

29-JUL-1999; 99US-0146447P.

(UROG-) UROGENESYS INC.

(AFAR/) AFAR D E.

(HUBER/) HUBERT R S.

(MITC/) MITCHELL S C.

Afar DE, Hubert RS, Mitchell SC;

WPI, 2000-303772/26.

Novel putative transcription factor gene 36P1A6 for treatment, diagnosis

and prevention of prostate, bladder, cervical, ovarian, pancreatic, and

colonic cancer.

Example 1; Page 29; 53pp; English.

The human 36P1A6 gene encodes a putative transcription factor based on

homology to the murine EHF gene which encodes a transcription factor

which is a member of the ETS family. 36P1A6 is expressed in androgen-

dependent and androgen-independent LAPC prostate cancer xenografts and in

normal prostate at approximately equal levels. The highest expression is

in the prostate and colon. 36P1A6 may be involved in activating tumor-

promoting genes or repressing genes that block tumorigenesis. The 36P1A6

polynucleotides and polypeptides are used for the treatment and diagnosis

of cancer, e.g. prostate, bladder, cervical, ovarian, pancreatic and

colonic cancer (all claimed). Anti-36P1A6 antibodies may be used for

purifying 36P1A6 and for isolating 36P1A6 homologues. Antisense

oligonucleotides and ribozymes can be used to inhibit the transcription

and translation of the 36P1A6 gene (claimed). The 36P1A6 polynucleotides

and polypeptides and immunogenic fragments may also be used in cancer

vaccines (claimed)

Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 3; Length 43;

Best Local Similarity 91.3%; Pred. No. 7.3e+02; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAACTAAGCTTGATC 23

27 AAAAAAAAAAAAAAAAAAGCTTGATC 5

RESULT 35

AAD06225/C

AAD06225;

31-JUL-2001 (first entry)

Human SGP28 CDNA synthesis primer, DPNCDN.

Human; specific granule protein 28; SGP28; therapy; primer; prostate;

colon; cancer; prognosis; vaccine; anticancer; ss.

Homo sapiens.

WO200131343-A2.

03-MAY-2001.

27-OCT-2000; 2000WO-US029607.

28-OCT-1999; 99US-0162610P.

(UROG-) UROGENESYS INC.

Hubert RS, Raitano AB, Afar DEH, Mitchell SC, Faris M;

Jakovovite A;

WPI; 2001-308685/32.

Detecting cancers, particularly of prostate and colon, from

overexpression of SGP28 protein, also methods for treating these cancers

e.g. by vaccination with the protein.

Example 1; Page 58; 102pp; English.

The present invention relates to methods and compositions for the

diagnosis and therapy of prostate cancer which utilizes human SGP28

(specific granule protein 28) gene and proteins. The method involves

detecting cancers, particularly of prostate and colon, from

overexpression of SGP28 protein. The expression of SGP28, which is an

extracellular protein is restricted to the prostate and ovary, and is

markedly up-regulated in prostate tumours. SGP28 sequence is used for

diagnosis (including in vivo imaging), staging, monitoring and prognosis

of prostate and colon cancer, and for assisting selection of therapy.

Also SGP28-expressing cancers can be treated by administering a

composition or vaccine that contains a vector expressing an antibody

specific for SGP28 protein, nucleic acid encoding SGP28 protein or its

fragments, polypeptides encoded by SGP28 gene and SGP28-specific antibody

optionally conjugated to toxin or therapeutic agent. SGP28 gene product

is also used as source of therapeutic antisense or ribozyme agents, as

primer/probes for diagnosis or prognosis, to identify compounds that

inhibit calcium entry into prostatic cells, for recombinant production of

SGP28 peptides and for isolating related sequences. SGP28 protein and its

fragments are used to raise specific antibodies (Ab) and to identify

specific binding agents (potentially useful as therapeutic and diagnostic

agents) and also potential anticancer agents. The present sequence is a

DPNCDN primer used to synthesize human SGP28 cDNA

Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 4; Length 43;

Best Local Similarity 91.3%; Pred. No. 7.3e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 AAAAAAAAAAAGCTTGATC 23
   |||||
DB 27 AAAAAAAAAAAGCTTGATC 5

RESULT 36
ID AAD04804 standard; DNA; 43 BP.
XX
XX AAD04804;
AC AAD04804;
XX
XX 17-JUN-2001 (first entry)
XX
XX Human secreted tumour antigen 36p65 cDNA synthesising primer, DPNCDN.
DE
XX Human, 36p65 protein; secreted tumour antigen; therapy; cancer; kidney;
KM bladder; ovary; breast; pancreas; colon; lung; vaccine; cytostatic;
KM primer; ss.
XX
XX Homo sapiens.
OS
XX WO200131015-A2.
XX
XX 03-MAY-2001.
XX
XX 30-OCT-2000; 2000WO-US029894.
XX
XX 28-OCT-1999; 99US-0162417P.
XX
XX (UROC-) UROGENESYS INC.
XX
XX Raitano AB, Jakobovits A, Paris M, Afar DEH, Hubert RS;
PI Mitchell SC;
XX
XX WPI; 2001-308646/32.
XX
XX Detecting presence of cancer expressing 36p65 protein in individual by
PT comparing protein level in test sample to normal sample, where elevated
PT level of protein in test sample indicates presence of cancer.
XX
XX Example 1; Page 69, 113pp; English.
XX
XX The present invention relates to a gene and its encoded secreted tumour
CC antigen, termed 36p65. These sequences are used for the diagnosis and
CC treatment of various cancers which express 36p65, such as cancers of the
CC kidney, bladder, ovary, breast, pancreas, colon and lungs. In normal
CC individuals 36p65 protein, is predominantly expressed in pancreas, with
CC lower levels of expression in prostate and small intestine. Vaccines
CC comprising immunogenic protein of 36p65 is useful for inhibiting the
CC development of prostate or colon cancer. Pharmaceutical composition
CC comprising 36p65 protein is useful for diagnosis and/or prognosis of
CC prostate cancer and other cancers, for modulating or inhibiting the
CC expression of 36p65 genes and/or translation of the 36p65 transcripts,
CC and as therapeutic agents. The present sequence is human 36p65 cDNA
CC synthesising primer, DPNCDN. Note: This sequence SEQ.ID.NO.3 is stated as
CC being similar to the sequence shown in sequence listing. However this
CC sequence contains 29 additional T bases at 3' end, which is missing in
CC the sequence shown in sequence listing.
XX
XX Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
SQ

Query Match 79.2%; Score 19.8; DB 4; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAGCTTGATC 23
   |||||
DB 27 AAAAAAAAAAAGCTTGATC 5

RESULT 37
AAF76005/c
ID AAF76005 standard; DNA; 43 BP.

```

```

XX
XX AAF76005;
AC
XX 22-MAY-2001 (first entry)
XX
XX RT primer, SEQ ID NO:11, used in human PC-L ECTIN cDNA isolation.
DE
XX
XX Human; PC-L ECTIN; C-type lectin; transmembrane antigen; normal testis;
KM layilin homologue; prostate cancer antigen; overexpression;
KM androgen-dependent prostate cancer; diagnosis; prognosis;
KM reverse transcription; RT primer; ss.
XX
XX Synthetic.
OS
XX WO200112811-A1.
XX
XX 22-FEB-2001.
XX
XX 11-AUG-2000; 2000WO-US022065.
XX
XX 12-AUG-1999; 99US-0148935P.
XX
XX (UROC-) UROGENESYS INC.
XX
XX Afar DEH, Hubert RS, Jakobovits A, Raitano AB;
PI WPI; 2001-211222/21.
XX
XX New PC-L ECTIN polynucleotide encoding a transmembrane antigen over
PT expressed in human prostate cancer, useful for the prognosis, diagnosis
PT and treatment of prostate cancer.
XX
XX Example 1; Page 58, 116pp; English.
XX
XX The invention relates to a novel human C-type lectin transmembrane
CC antigen, PC-L ECTIN (AAB73309) and cDNA encoding it (AAF76004). The
CC expression of the human PC-L ECTIN gene is normally restricted to the
CC testis, but is highly overexpressed in prostate cancer. PC-L ECTIN
CC expression is higher in androgen-dependent prostate tumours compared with
CC androgen-independent prostate tumours, and expression is therefore likely
CC to be dependent on the presence of androgen. Human PC-L ECTIN therefore
CC represents a diagnostic and therapeutic target for prostate cancer.
CC particularly androgen-dependent prostate cancer. Human PC-L ECTIN exhibits
CC homology to hamster layilin (44.9% identity over a 265 residue overlap),
CC but is not thought to be the human orthologue of layilin, as diverges
CC significantly in a key functional domain proposed for the layilin
CC protein. Human PC-L ECTIN or an immunogenic portion thereof, a vector
CC encoding PC-L ECTIN, a PC-L ECTIN antisense nucleotide, a PC-L ECTIN
CC nucleotide-targeted ribozyme, or an anti- PC-L ECTIN antibody may be used
CC to prepare a composition for treating a patient with a cancer,
CC particularly prostate cancer, but also breast, bladder, lung, bone,
CC colon, pancreatic, testicular, cervical or ovarian cancers that express
CC PC-L ECTIN. PC-L ECTIN proteins are also useful for diagnosing the presence
CC of cancer. PC-L ECTIN antibodies and nucleotides are useful in the
CC treatment (e.g., antisense therapy), diagnosis and/or prognosis of
CC prostate cancer and other PC-L ECTIN-expressing cancers. PC-L ECTIN
CC antibodies may also be used as drug targeting agents. The PC-L ECTIN
CC nucleotides and proteins may additionally be used in drug discovery to
CC identify molecules that modulate PC-L ECTIN function or expression. The
CC present sequence represents a reverse transcription (RT) primer used in
CC the isolation of human PC-L ECTIN cDNA
XX
XX Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
SQ

Query Match 79.2%; Score 19.8; DB 4; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAGCTTGATC 23
   |||||
DB 27 AAAAAAAAAAAGCTTGATC 5

```

```
RESULT 38
ABL50400/c
ID ABL50400 standard; DNA; 43 BP.
XX
AC ABL50400;
XX
XX 17-JUN-2002 (first entry)
DT
XX Human 158P1H4 cDNA synthesis primer DPNCDN SEQ ID NO:717.
DE
XX Human; 158P1H4; chromosome 8q220q23, 158P1F4; chromosome 8q23; cancer;
KW bladder cancer; immune response; cytotoxic T lymphocyte; CTL; HLA;
KW human leukocyte antigen; helper T lymphocyte; HTL; PCR primer; adapter;
KW ss.
XX Homo sapiens.
OS Synthetic.
XX WO200216598-A2.
XX
XX 28-FEB-2002.
XX
XX 22-AUG-2001; 2001WO-US026411.
XX
XX 22-AUG-2000; 2000US-0227098P.
XX
XX 10-APR-2001; 2001US-0282739P.
XX
XX (AGEN-) AGENSYS INC.
XX
XX Chailita-Bid PM, Hubert RS, Raitano AB, Afar DEH, Levin E;
XX Paris M, Ge W, Jakobovits A;
XX WPI; 2002-269357/31.
XX
XX Monitoring 158P1H4 gene products in biological sample from patient who
PT has or is suspected of having cancer, useful for treating cancer,
PT comprises identifying presence of aberrant 158P1H4 gene products in
PT biological sample.
XX
XX Example 1; Page 69; 209pp; English.
XX
XX The present invention describes a method for monitoring 158P1H4 gene
CC products in a biological sample from a patient who has or is suspected of
CC having cancer. The method comprises determining the status of 158P1H4
CC gene products in a tissue sample from an individual, comparing the status
CC to the status of 158P1H4 gene products in a normal sample, and
CC identifying the presence of aberrant 158P1H4 gene products in the sample.
CC 158P1H4 sequences have cytostatic activity and can be used in vaccine
CC production. 158P1H4 polynucleotides may be used in monitoring genetic
CC abnormalities. The 158P1H4 proteins may be used in assessing the status
CC of 158P1H4 gene products in normal versus cancerous tissues and so
CC elucidating the malignant phenotype, in generating and characterizing
CC domain-specific antibodies, for identifying agents or cellular factors
CC that bind to 158P1H4 or its particular domain, and for generating cancer
CC vaccines. Antibodies against 158P1H4 are useful in diagnostic and
CC prognostic assays, in treating patients with cancer, in generating
CC cytotoxic T lymphocyte (CTL) or helper T lymphocyte (HTL) responses, and
CC as immunological reagents for detecting 158P1H4-expressing cells. The
CC antibodies are particularly useful in bladder cancer diagnostic and
CC prognostic assays, and imaging methodologies. The 158P1H4 gene has been
CC located to chromosome 8q22-q23, and the 158P1F4 gene also described in
CC ABL50429 and ABB94468 to ABB95188 represent sequences used in the
CC exemplification of the present invention
XX
XX Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 6; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
DB 27 AAAAAAAAAAAAGCTTGATC 5
RESULT 39
ABL50412/c
ID ABL50412 standard; DNA; 43 BP.
XX
AC ABL50412;
XX
XX 17-JUN-2002 (first entry)
DT
XX Human 158P1F4 cDNA synthesis primer DPNCDN SEQ ID NO:729.
DE
XX Human; 158P1H4; chromosome 8q220q23, 158P1F4; chromosome 8q23; cancer;
KW bladder cancer; immune response; cytotoxic T lymphocyte; CTL; HLA;
KW human leukocyte antigen; helper T lymphocyte; HTL; PCR primer; adapter;
KW ss.
XX Homo sapiens.
OS Synthetic.
XX WO200216598-A2.
XX
XX 28-FEB-2002.
XX
XX 22-AUG-2001; 2001WO-US026411.
XX
XX 22-AUG-2000; 2000US-0227098P.
XX
XX 10-APR-2001; 2001US-0282739P.
XX
XX (AGEN-) AGENSYS INC.
XX
XX Chailita-Bid PM, Hubert RS, Raitano AB, Afar DEH, Levin E;
XX Paris M, Ge W, Jakobovits A;
XX WPI; 2002-269357/31.
XX
XX Monitoring 158P1H4 gene products in biological sample from patient who
PT has or is suspected of having cancer, useful for treating cancer,
PT comprises identifying presence of aberrant 158P1H4 gene products in
PT biological sample.
XX
XX Example 45; Page 116; 209pp; English.
XX
XX The present invention describes a method for monitoring 158P1H4 gene
CC products in a biological sample from a patient who has or is suspected of
CC having cancer. The method comprises determining the status of 158P1H4
CC gene products in a tissue sample from an individual, comparing the status
CC to the status of 158P1H4 gene products in a normal sample, and
CC identifying the presence of aberrant 158P1H4 gene products in the sample.
CC 158P1H4 sequences have cytostatic activity and can be used in vaccine
CC production. 158P1H4 polynucleotides may be used in monitoring genetic
CC abnormalities. The 158P1H4 proteins may be used in assessing the status
CC of 158P1H4 gene products in normal versus cancerous tissues and so
CC elucidating the malignant phenotype, in generating and characterizing
CC domain-specific antibodies, for identifying agents or cellular factors
CC that bind to 158P1H4 or its particular domain, and for generating cancer
CC vaccines. Antibodies against 158P1H4 are useful in diagnostic and
CC prognostic assays, in treating patients with cancer, in generating
CC cytotoxic T lymphocyte (CTL) or helper T lymphocyte (HTL) responses, and
CC as immunological reagents for detecting 158P1H4-expressing cells. The
CC antibodies are particularly useful in bladder cancer diagnostic and
CC prognostic assays, and imaging methodologies. The 158P1H4 gene has been
CC located to chromosome 8q22-q23, and the 158P1F4 gene also described in
CC ABL50429 and ABB94468 to ABB95188 represent sequences used in the
CC exemplification of the present invention
XX
XX Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 6; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

OY 1 AAAAAAAAACTAAGCTTGATC 23
 Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5

RESULT 40
 ABA03602/c
 ID ABA03602 standard; DNA; 43 BP.
 XX
 AC ABA03602;
 XX
 DT 08-FEB-2002 (first entry)
 XX
 DE Human cDNA synthesis primer DPNCDN.
 XX

Human, 343PD7; cytoskeletal; vaccine; gene therapy; cancer;
 human leukocyte antigen; HLA; major histocompatibility complex; MHC;
 HLA A1; HLA A11; HLA A02; HLA A24; HLA A3; HLA B35; HLA B7;
 cDNA synthesis; primer; ss.

XX Homo sapiens.
 XX WO200159110-A2.
 XX 16-AUG-2001.
 XX PD 08-FEB-2001; 2001WO-US004094.
 XX PE 08-FEB-2001; 2000US-0181020P.
 XX PR 08-FEB-2000; 2000US-0181020P.
 XX

PA (UROG-) UROGENESYS INC.
 PI Faris M, Afar DEH, Challita-Eid PM, Hubert RS, Levin E;
 PI Mitchell SC, Jakobovits A;
 XX WPI; 2002-025689/03.
 XX

PT New gene designated 34PD7, encoding a tissue-specific protein highly
 PT expressed in prostate cancer, for use as diagnostic and/or therapeutic
 PT target for cancers, and for eliciting an immune response.
 XX

PS Example 1; Page 53; 112pp; English.

CC The invention relates to a polynucleotide, designated 34PD7, encoding a
 CC 34PD7-related protein, comprising a sequence of 2198 nucleotides fully
 CC defined in the specification. The presence of elevated 34PD7 mRNA or
 CC protein expression indicates the presence of cancer occurring in
 CC prostate, bladder, kidney, bone, cervical, uterine, ovarian,
 CC breast, pancreatic, stomach, colon, rectal leukocytes, liver, and lung
 CC tissue, and in melanocytes. An antibody against the 34PD7-related
 CC protein, an antisense polynucleotide complementary to 34PD7
 CC polynucleotide, or a ribozyme capable of cleaving the 34PD7
 CC polynucleotide is useful for inhibiting the development of a cancer
 CC expressing 34PD7 in a patient. The present sequence was used in an
 CC example demonstrating suppression subtractive hybridisation (SSH) -
 CC generated isolation of a cDNA fragment of the 34PD7 gene
 XX

SO Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 43;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAGCTTGATC 23
 Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5

RESULT 41
 AAS99436/c
 ID AAS99436 standard; DNA; 43 BP.
 XX

AC AAS99436;
 XX 12-MAR-2002 (first entry)
 DT DPNCDN cDNA synthesis primer.
 DE
 XX
 XX Human; 98P7C3; ss; homeodomain protein; vaccine; cytotoxic; epitope;
 KW transgenic animal; immunogen; T cell; cytotoxic T cell; CTL;
 KW prostate cancer; bladder cancer; kidney cancer; lung cancer;
 KW breast cancer; uterine cancer; cervical cancer; stomach cancer;
 KW rectal cancer; colon cancer; chromosome 4q11-q12; PCR primer; adapter;
 KW suppression subtractive hybridisation; SSH.
 XX

XX Synthetic.
 XX WO200190157-A2.
 XX 29-NOV-2001.
 XX PD 24-MAY-2001; 2001WO-US017495.
 XX PE 24-MAY-2000; 2000US-0207138P.
 XX PR 24-MAY-2000; 2000US-0207138P.
 XX

PA (UROG-) UROGENESYS INC.
 PI Challita-Eid PM, Hubert RS, Faris M, Afar DEH, Levin E;
 PI Mitchell SC, Jakobovits A;
 XX WPI; 2002-097642/13.
 XX

PT New isolated 98P7C3-related homeodomain protein highly expressed in
 PT various cancers, useful in cancer vaccines and for generating immune
 PT response directed to 98P7C3 in mammal.
 XX

PS Example 1; Page 53; 155pp; English.

CC The invention relates to an isolated 98P7C3-related protein which is a
 CC homeodomain protein highly expressed in various cancers. Also include are
 CC polynucleotides encoding the protein or proteins 90% identical to 98P7C3,
 CC a pharmaceutical composition comprising the polynucleotides (including an
 CC expression vector comprising the 98P7C3 encoding polynucleotides) or a
 CC host cell transformed with the vector, an anti-98P7C3 antibody, a non-
 CC human transgenic animal expressing a 98P7C3 protein, methods of detecting
 CC the 98P7C3 protein or polynucleotides in a biological sample, monitoring
 CC the presence of cancer in an individual by detecting an elevated level of
 CC the 98P7C3 protein or polynucleotides and a pharmaceutical composition
 CC comprising a modulator of 98P7C3, 98P7C3 protein, or T cell/B cell
 CC epitopes derived from it, are useful in inducing an immune response (in
 CC mammal) to a 98P7C3 protein. Upon contact with a cytotoxic T cell (CTL)
 CC the immunogens induce the CTL (with its helper T cell) to kill an
 CC autologous cell expressing 98P7C3. The immunogen may be a nucleic acid
 CC encoding the protein or epitope. The antibody is useful for delivering a
 CC cytotoxic agent to a cell that expresses 98P7C3, by conjugating the
 CC cytotoxic agent to the antibody or its fragment that specifically binds
 CC to a 98P7C3 epitope, and exposing the cell to the antibody-agent
 CC conjugate. The modulator is useful for treating a patient with a cancer
 CC that expresses 98P7C3 (e.g. prostate cancer, bladder cancer, kidney
 CC cancer, lung cancer, breast cancer, uterine cancer, cervical cancer,
 CC stomach cancer, rectal cancer and colon cancer), by administering to the
 CC patient a vector that comprises the modulator, such that the vector
 CC delivers a single chain monoclonal antibody coding sequence to the cancer
 CC cells and the encoded single chain antibody is expressed intracellularly
 CC in it. The gene for 98P7C3 is located on human chromosome 4q11-q12. The
 CC present sequence is oligonucleotide adapter or PCR primer used to isolate
 CC a cDNA sequence for 98P7C3 by the method of suppression subtractive
 CC hybridisation, SSH
 XX

SO Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 43;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTTGATC 23
 |||||
 Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5

RESULT 42
 ABK67415/C
 ID ABK67415 standard; DNA; 43 BP.
 AC ABK67415;
 XX
 DT 02-JUN-2002 (first entry)
 XX

Human 83P2H3 cDNA isolation primer DPNCDN.
 DE
 KW Human; human leukocyte antigen; HLA; immunogen; 83P2H3; CatrF2B11;
 KW calcium transport protein; cancer; prostate cancer; cytostatic;
 KW chromosome 7q34; chromosome 12q24.1; T cell; B cell; ss; primer.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS
 PN WO200214361-A2.
 XX
 PD 21-FEB-2002.
 XX
 PF 17-AUG-2001; 2001WO-US025782.
 XX
 PR 17-AUG-2000; 2000US-0226325P.
 XX
 PA (AGEN-) AGENSYS INC.
 XX
 PI Raitano AB, Challita-Bid PM, Faria M, Saffran DC, Afar DEH;
 PI Levin E, Hubert RS, Ge W, Jakobovits A;
 XX
 DR WPI: 2002-269179/31.

Monitoring 83P2H3 gene products for monitoring the presence of cancer in a subject, comprises determining the status of 83P2H3 gene products in a tissue sample from the subject and comparing it to a normal sample.

Example 1; Page 75; 270pp; English.

The invention relates to monitoring 83P2H3 (a calcium transport protein whose gene is located on chromosome 7q34) gene products in a biological sample from a patient who has or is suspected of having cancer (especially prostate cancer), comprises: (a) determining the status of 83P2H3 gene products expressed by cells in a tissue sample from an individual and (b) comparing the status to the status of 83P2H3 gene products in a normal sample. Also included are modulators of 83P2H3 function or status, generating antibodies/immune response against 83P2H3 (or related protein CatrF2B11 whose gene is located on chromosome 12q24.1) using identified HLA (human leukocyte antigen) binding peptides derived from the protein, delivering a cytotoxic agent to a cell expressing 83P2H3 by conjugating the agent to an anti-83P2H3 antibody, a recombinant protein comprising an antigen-binding region of the antibody, a non-human transgenic animal that produces the recombinant protein, a hybridoma that produces the recombinant protein, a single-chain monoclonal antibody that comprises the variable domains of the heavy and light chains of the anti-83P2H3 antibody, a vector comprising a polynucleotide that encodes the monoclonal antibody and inducing an immune response to a 83P2H3 protein, by providing a 83P2H3-related protein that comprises a T cell or B cell epitope, and contacting the epitope with an immune system T cell or B cell, respectively. The method is useful for monitoring 83P2H3 gene products in a biological sample for monitoring the presence of cancer in an individual. The modulator is useful for inhibiting the growth of cancer cells that express 83P2H3, for treating cancer and the vector is useful for treating a patient with a cancer that expresses 83P2H3. The immunological methods are useful for generating an immune response against 83P2H3, and for detecting the presence of 83P2H3-related protein or polynucleotide in a biological sample from a patient who has or who is suspected of having cancer. The antibody is useful in prostate cancer diagnosis, prognosis, imaging

CC methodologies and treatment, to detect and quantify 83P2H3 and mutant
 CC 83P2H3-related proteins, for purifying a 83P2H3-related protein, for
 CC isolating 83P2H3 homologues/related molecules, and for generating anti-
 CC idiotypic antibodies that mimic the 83P2H3 protein. The present sequence
 CC is a PCR primer used in the isolation of cDNA encoding 83P2H3 or its
 CC related protein CatrF2B11
 CC
 XX

SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 43;
 Best Local Similarity 91.3%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTTGATC 23
 |||||
 Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5

RESULT 43
 AAL53469/C
 ID AAL53469 standard; DNA; 43 BP.
 AC AAL53469;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DT
 XX
 DE Zinc transporter protein 108P5H8 gene PCR primer DPNCDN.
 XX
 KW Cytostatic; gene therapy; vaccine; zinc transporter protein 108P5H8;
 KW cancer; breast; colon; ovarian; lung; humoral; cellular immune response;
 KW passive immunisation; PCR; primer; ss.
 XX
 OS Unidentified.
 OS
 PN WO200260953-A2.
 XX
 PD 08-AUG-2002.
 XX
 PF 17-DEC-2001; 2001WO-US049133.
 XX
 PR 15-DEC-2000; 2000US-0256210P.
 XX
 PA (AGEN-) AGENSYS INC.
 XX
 PI Challita-Bid PM, Faria M, Afar DEH, Hubert RS, Mitchell SC;
 PI Levin E, Morrison KM, Raitano AB, Jakobovits A;
 XX
 DR WPI: 2002-627469/67.

Composition comprising a substance that modulates the status of a zinc transporter protein (108P5H8), useful in diagnosing and treating patients with cancer that expresses 108P5H8, such as breast, colon, ovarian or lung cancer.

Example 1; Page 94; 309pp; English.

The invention relates to a new composition comprising a substance that modulates the status of a zinc transporter protein, designated as 108P5H8, or a molecule that is modulated by 108P5H8. The composition is useful in diagnosing, preventing, prognosticating or treating patients with cancer that expresses 108P5H8, such as breast, colon, ovarian or lung cancer. The 108P5H8 gene or its fragment can be used to elicit a humoral or cellular immune response. The antibodies are useful in active or passive immunisation. The 108P5H8 polynucleotides are useful as probes and primers for the amplification or detection of 108P5H8 genes, as coding sequences for directing the expression of 108P5H8 polypeptides, or as tools for modulating or inhibiting the expression of 108P5H8 genes. The polynucleotides of the invention can be used to treat disorders by gene therapy. This polynucleotide sequence represents a zinc transporter protein 108P5H8 related PCR primer of the invention

Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

```

RESULT 44
ABZ78169/C
ID ABZ78169 standard; DNA; 43 BP.
XX
XX AC ABZ78169;
XX DT 19-MAY-2003 (first entry)
XX DE cDNA synthesis primer, DPNCND.
XX KM Cytostatic; vaccine; cancer; immune response; PCR; primer; ss.
XX OS Synthetic.
XX PN WO200283921-A2.
XX PD 24-OCT-2002.
XX PF 10-APR-2002; 2002MO-USO11654.
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 10-APR-2001; 2001US-0283112P.
XX PR 25-APR-2001; 2001US-0286630P.
XX PA (AGEN-) AGENSYS INC.
XX PI Jakobovits A, Challita-Eld PM, Faris M, Ge W, Hubert RS;
XX PI Morrison K, Morrison RK, Raitano AB;
XX DR MPI; 2003-075555/07.
XX
XX PT New composition comprising a substance that modulates the structure of
PT proteins and polynucleotides, useful for therapeutic, prognostic and
PT diagnostic reagents for eliciting cellular or humoral immune response in
PT cancer patients.
XX
XX PS Example 1; Page 72; 1021pp; English.
XX
CC The present invention relates to novel human cancer-related genes and
CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and
CC proteins are useful for eliciting a humoral or cellular immune response.
CC The genes are useful as probes and primers for the amplification and/or
CC detection of genes, mRNAs or their fragments, as reagents for the
CC diagnosis and/or prognosis of cancer, as coding sequences capable of
CC directing the expression of the protein, as tools for modulating or
CC inhibiting the expression of genes and/or translation of transcripts, and
CC as therapeutic agents. The proteins and peptides are useful as
CC therapeutic, prognostic and diagnostic reagents for cancer. The present
CC sequence is a primer, used in an example from the invention
XX
SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match          79.2%; Score 19.8; DB 8; Length 43;
Best Local Similarity 91.3%; Pred.No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      1 AAAAAAAAAAACTTAAAGCTTGATC 23
Db      27 AAAAAAAAAAAAAAAGCTTGATC 5

```

```

XX AC ABZ20556;
XX XX
XX DT 03-MAR-2003 (first entry)
XX DE Cancer associated coding sequence cDNA synthesis primer.
XX XX
XX KW Cancer associated coding sequence; cancer; human; cytostatic;
XX gene therapy; PCR; primer; ss.
XX OS Homo sapiens.
XX PN WO200283920-A2.
XX PD 24-OCT-2002.
XX PF 10-APR-2002; 2002WO-US011645.
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 10-APR-2001; 2001US-0283112P.
XX PR 25-APR-2001; 2001US-0286630P.
XX PR 10-APR-2002; 2002US-00286630.
XX PA (AGEN-) AGENSYS INC.
XX PI Jakobovits A, Hubert RS, Challita-Eld PM;
XX DR WPI; 2003-093030/08.
XX PT New pharmaceutical composition for diagnosing, prognosing, preventing or
XX PT treating cancer; comprises a substance that modulates a nucleic acid
XX PT sequence, e.g. 105P1B7, 152P1A2B or 156P3A6, or a molecule modulated by
XX PT the nucleic acid.
XX PS
XX PS Example 1; Page 33; 72pp; English.
XX CC
XX CC The present invention relates to a pharmaceutical composition comprising
XX CC a substance that modulates the status of a cancer associated nucleic acid
XX CC sequence such as given in the specification (see ABZ20564-ABZ20575) or a
XX CC molecule that is modulated by the above nucleic acid sequence, where the
XX CC status of a cell that expresses the nucleic acid sequence is modulated.
XX CC The composition is useful in diagnosing, prognosing, preventing and/or
XX CC treating cancer. The nucleic acid sequence may be used in monitoring
XX CC genetic abnormalities, in generating and characterising domain-specific
XX CC antibodies, for identifying agents or cellular factors that bind to a
XX CC protein, and in therapeutic and diagnostic contexts, such as diagnostic
XX CC assay, cancer vaccine, and methods of preparing vaccines. The present
XX CC sequence is a primer used to identify the cancer associated coding
XX CC sequences suitable to be modulated in the method of the invention
XX SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
XX
XX Query Match 79.2%; Score 19.8; DB 8; Length 43;
XX Best Local Similarity 91.3%; Pred. No. 7.3e+02;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX Qy 1 AAAAAAAAACTAAGCTTGATC 23
XX Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5
XX
XX RESULT 46
XX ID ADCT1176/c
XX AC ADCT1176 standard; DNA; 43 BP.
XX ADCT1176;
XX DT 18-DEC-2003 (first entry)
XX DE cDNA synthesis primer identified as DPNCND.
XX KW 205P1B5; prostate cancer; immune response; transgenic; knock out animal;
XX cytostatic; immunogenic; vaccine; ss; DPNCND; primer; PCR.

```

```

XX OS Unidentified.
XX XX WO2003020954-A2.
XX XX 13-MAR-2003.
XX PF 30-AUG-2002; 2002WO-US027760.
XX PR 31-AUG-2001; 2001US-0316664P.
XX PA (AGEN-) AGENSYS INC.
XX PI Chailita-Bid PM, Raitano AB, Faris M, Hubert RS, Jakobovits A;
XX DR WPI; 2003-354484/33.
XX PT New polynucleotide designated 205PIB5, for diagnosing and treating
XX PT prostate cancer, and as probes or primers for the amplification and/or
XX PT detection of 205PIB5 genes.
XX Example 1; SEQ ID NO 714; 162bp; English.
XX PS
XX CC This invention relates to a novel gene designated 205PIB5, and the
XX CC encoded protein, which is aberrantly expressed in prostate cancer.
XX CC Specifically, it refers to the two variants of 205PIB5 mapped to
XX CC chromosome 8p21-8p12, namely 205PIB5v1 and 205PIB5v2 and fragments
XX CC thereof that serve as useful diagnostic, prophylactic, prognostic and/or
XX CC therapeutic targets for prostate and other types of cancers. The present
XX CC invention describes methods for the isolation of 205PIB5, for generating
XX CC an immune response and for generating transgenic or knock out animals for
XX CC the development and screening of therapeutically useful reagents.
XX CC Furthermore, it refers to identifying proteins, small molecules or other
XX CC agents that interact with 205PIB5, and can be used to identify pathways
XX CC activated by 205PIB5. Accordingly, these are cytostatic and immunogenic
XX CC compositions that are useful for the development of cancer vaccines. This
XX CC oligonucleotide sequence is the cDNA synthesis primer designated DPNCND
XX CC used in an exemplification of the invention.
XX SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
XX
XX Query Match          79.2%; Score 19.8; DB 10; Length 43;
XX Best Local Similarity 91.3%; Pred. No. 7.3e+02;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAACTAAAGCTTGATC 23
XX      |||||
XX Db 27 AAAAAAAAAAAAGCTTGATC 5
XX
XX RESULT 47
XX ADD84526/C
XX ID ADD84526 standard; DNA; 43 BP.
XX AC ADD84526;
XX XX 29-JAN-2004 (first entry)
XX DT
XX DE DPNCND cDNA synthesis primer SEQ ID NO:714.
XX XX
XX KM 121PIF1; 121PIF1 modulation; human; chromosome 4q; cytostatic;
XX KM gene therapy; vaccine; cancer; immune response; immunisation; primer; ss.
XX OS Synthetic.
XX PN WO200295009-A2.
XX PD 28-NOV-2002.
XX XX 28-FEB-2002; 2002WO-US006242.
XX PF
XX PR 05-MAR-2001; 2001US-00799250.
XX DR
XX

```

```

PA (AGEN-) AGENSYS INC.
XX XX Chailita-Bid PM, Hubert RS, Raitano AB, Faris M, Afar DEH, Ge W;
XX PI Jakobovits A;
XX XX WPI; 2003-156757/15.
XX DR
XX PT Composition comprising a substance that modulates the status of 121PIF1,
XX PT useful in diagnosing, preventing, prognosticating or treating patients
XX PT with cancer that expresses 121PIF1, such as breast, colon, ovarian or
XX PT lung cancer.
XX Example 1; Page 71; 285bp; English.
XX PS
XX CC The present invention describes a composition (I) comprising a substance
XX CC that modulates the status of 121PIF1 (gene and encoded protein), or a
XX CC molecule that is modulated by 121PIF1, where the status of a cell that
XX CC expresses 121PIF1 is modulated. The human 121PIF1 gene maps to chromosome
XX CC 4q. (I) has cytostatic activity, and can be used in gene therapy, and in
XX CC vaccines. The composition (I) can be used for diagnosing, preventing,
XX CC prognosticating or treating patients with cancer that expresses 121PIF1,
XX CC such as breast, colon, ovarian or lung cancer. The 121PIF1 gene or its
XX CC fragment can be used to elicit a humoral or cellular immune response.
XX CC 121PIF1 antibodies can be used in active or passive immunisation. 121PIF1
XX CC polynucleotides are useful as probes and primers for the amplification or
XX CC detection of 121PIF1 genes, as coding sequences for directing the
XX CC expression of 121PIF1 polypeptides, or as tools for modulating or
XX CC inhibiting the expression of 121PIF1 genes. The present sequence is used
XX CC in the exemplification of the present invention.
XX SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
XX
XX Query Match          79.2%; Score 19.8; DB 10; Length 43;
XX Best Local Similarity 91.3%; Pred. No. 7.3e+02;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAACTAAAGCTTGATC 23
XX      |||||
XX Db 27 AAAAAAAAAAAAGCTTGATC 5
XX
XX RESULT 48
XX ABV99869/C
XX ID ABV99869 standard; DNA; 43 BP.
XX AC ABV99869;
XX XX 28-MAR-2003 (first entry)
XX DT
XX DE Human 121P2A3 cDNA PCR synthesis primer DPNCND.
XX XX
XX KM Human; 121P2A3; cytostatic; immunostimulant; vaccine; PCR;
XX KM humoral immune response; cellular immune response; primer; ss.
XX OS Homo sapiens.
XX PN WO200283068-A2.
XX PD 24-OCT-2002.
XX XX 09-APR-2002; 2002WO-US011359.
XX PF
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 25-APR-2001; 2001US-0286630P.
XX PR 22-JUN-2001; 2001US-0300373P.
XX XX
XX PA (AGEN-) AGENSYS INC.
XX XX Chailita-Bid PM, Raitano AB, Faris M, Hubert RS, Mitchell SC;
XX PI Afar DEH, Safifan D, Morrison K, Morrison RK, Ge W, Jakobovits A;
XX DR WPI; 2003-092956/08.
XX

```

PT New composition comprising a substance that modulates the status of
PT 121P2A3 polypeptides, useful for eliciting humoral or cellular immune
PT responses or in assessing the status of 121P2A3 gene products in normal
PT versus cancerous tissues.
XX
PS Example 1; Page 70; 362pp; English.
XX
CC The invention relates to a novel composition comprising a substance that
CC modulates the status of a protein, 121P2A3. The composition of the
CC invention has cytostatic and immunostimulant activity, and is useful as a
CC vaccine. The 121P2A3 proteins and polynucleotides are useful for
CC eliciting humoral or cellular immune response. The polynucleotides are
CC useful for characterizing cytogenetic abnormalities of this chromosomal
CC locus, as tools that can be used to delineate cytogenetic abnormalities
CC in the chromosomal region that encodes 121P2A3 that may contribute to
CC malignant phenotype, and in assessing the status of 121P2A3 gene products
CC in normal versus cancerous tissues. The proteins are useful for
CC generating and characterizing domain-specific antibodies, for identifying
CC agents or cellular factors that bind to 121P2A3 or a particular structure
CC domain, and in various therapeutic and diagnostic contexts, including
CC cancer vaccines. The antibodies or T cells reactive with the product are
CC useful in passive or active immunisation, and in imaging methodologies
CC for the management of cancer. The present sequence represents the primer
CC DPNCDN, used in the invention to synthesise cDNA from RNA extracted from
CC xenograft tissue
XX
SO Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 10; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAACTAAAGCTTGATC 23
Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5
XX
RESULT 49
AD197735/c
ID AD197735 standard; DNA; 43 BP.
XX
AC AD197735;
XX
DT 04-NOV-2004 (first entry)
XX
DE 238P1B2 gene related DPNCDN cDNA synthesis primer.
XX
KW 238P1B2; transgenic animal; cytotoxic; cancer; cytostatic; gene therapy;
KW vaccine; prostate cancer; humoral; cellular immune response; primer; ss.
XX
OS Unidentified.
XX
FN WO2003085081-A2.
XX
PD 16-OCT-2003.
XX
PF 01-APR-2002; 2002WO-US010132.
XX
PR 01-APR-2002; 2002WO-US010132.
XX
PA (AGEN-) AGENSYS INC.
XX
PI Raitano AB, Challita-Bid PM, Faris M, Hubert RS, Morrison RK;
PI Ge W, Jakobovits A;
XX
DR WPI; 2003-812724/76.
XX
XX A composition useful for detecting, preventing or treating cancer (e.g.
PT prostate cancer) comprises a substance that modulates the status of
PT 238P1B2, or a molecule that is modulated by 238P1B2.
XX
PS Example 1; Page 68; 315pp; English.
XX

CC The invention relates to a novel composition comprising a substance that
CC modulates the status of 238P1B2, or a molecule that is modulated by
CC 238P1B2. The invention further relates to: a non-human transgenic animal
CC that produces an antibody; a hybridoma that produces an antibody;
CC delivering a cytotoxic agent or a diagnostic agent to a cell that
CC expresses 238P1B2; a polynucleotide that encodes a peptide; a
CC pharmaceutical composition that comprises the composition cited above, in
CC a human unit dosage form, where the substance comprises: a ribozyme that
CC cleaves a polynucleotide having 238P1B2 coding sequence or a nucleic acid
CC molecule that encodes the ribozyme, and a carrier; or human T-cells that
CC specifically recognise a 238P1B2 peptide sequence in the context of a
CC particular HLA molecule; inhibiting growth of cancer cells that express
CC 238P1B2; generating a mammalian immune response directed to 238P1B2; an
CC assay for detecting the presence of a 238P1B2-related protein or
CC polynucleotide in a biological sample; and monitoring the presence of
CC cancer in an individual. The novel composition has cytostatic activity.
CC The 238P1B2 polynucleotide can be used in gene therapy to treat
CC disorders. The 238P1B2 polypeptide can be used in the creation of a
CC vaccine. The composition and methods are useful in detecting, preventing,
CC prognosing or treating cancer (e.g. prostate cancer). The genes, proteins
CC or antibodies can be used to elicit a humoral or cellular immune
CC response. The polynucleotide may be used as a probe or a primer, or in
CC chromosomal mapping. This polynucleotide represents a primer used in the
CC exemplification of the invention.
XX
SO Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 11; Length 43;
Best Local Similarity 91.3%; Pred. No. 7.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAACTAAAGCTTGATC 23
Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5
XX
RESULT 50
ADK83217/c
ID ADK83217 standard; DNA; 43 BP.
XX
AC ADK83217;
XX
DT 20-MAY-2004 (first entry)
XX
DE 191P4D12(b) cDNA synthesis primer, SEQ ID 48.
XX
KW Cytostatic; Immunostimulant; Vaccine; Gene Therapy; 191P4D12(b); cancer;
KW prostate cancer; bladder cancer; kidney cancer; colon cancer;
KW lung cancer; pancreas cancer; ovary cancer; breast cancer; uterus cancer;
KW cervix cancer; immune response; human; chromosome 1q22-q23.2;
KW Suppression Subtractive Hybridisation; SSH; primer; ss.
XX
XX Synthetic.
XX
OS
XX
FN WO2004016799-A2.
XX
PD 26-FEB-2004.
XX
PF 23-APR-2003; 2003WO-US013013.
XX
PR 16-AUG-2002; 2002US-0404306P.
XX
PR 01-NOV-2002; 2002US-0423290P.
XX
PA (AGEN-) AGENSYS INC.
XX
PI Raitano AB, Challita-Bid PM, Jakobovits A, Faris M, Ge W;
PI WPI; 2004-203808/19.
XX
XX New composition comprising 191P4D12(b) proteins and polynucleotides,
PT useful in diagnosing, preventing and treating cancer, e.g. prostate,
PT bladder, kidney, colon, lung, pancreas or ovary and in eliciting a
PT humoral or cellular immune response.
XX

XX Example 1; SEQ ID NO 48; 443bp; English.

PS The present invention relates to novel compositions comprising peptides
 CC (ADK83300-ADK90584) from 191P4D12(b) and related proteins. It was found
 CC that 191P4D12(b), which maps to chromosome 1q22-q23.2, is aberrantly
 CC expressed in a number of cancers and so the compositions of the invention
 CC are useful in diagnosing, preventing and treating cancer, e.g. cancer of
 CC the prostate, bladder, kidney, colon, lung, pancreas, ovary, breast,
 CC uterus or cervix and in eliciting a humoral or cellular immune response.
 CC To isolate genes that are overexpressed in prostate cancer, the
 CC Suppression Subtractive Hybridisation (SSH) method was used, using cDNA
 CC derived from prostate cancer tissues. The present sequence was used in an
 CC example from the invention for SSH-generated isolation of cDNA fragments
 CC of the 191P4D12(b) gene.

XX SQ Sequence 43 BP; 3 A; 2 C; 2 G; 36 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 12; Length 43;

Best Local Similarity 91.3%; Pred. No. 7.3e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAACTAAGCTTGATC 23

Db 27 AAAAAAAAAAAAAAAAAAGCTTGATC 5

Search completed: December 14, 2005, 02:42:33
 Job time : 211.2 secs

THIS PAGE BLANK (USPTO)

C 96	20.2	80.8	558	3	B1500177	B1500177	1876f09.y
C 97	20.2	80.8	579	9	AQ273471	AQ273471	ndxb0030K
C 98	20.2	80.8	590	11	CR254381	CR254381	Forward s
C 99	20.2	80.8	603	2	BF621957	BF621957	HVSMa000
C 100	20.2	80.8	615	8	CX184687	CX184687	Al1.45-17
C 101	20.2	80.8	615	9	AZ861675	AZ861675	2M016814
C 102	20.2	80.8	615	10	CBE831786	CBE831786	t1gr-g88-
C 103	20.2	80.8	616	7	CO689848	CO689848	DD11-266d
C 104	20.2	80.8	618	1	AM640963	AM640963	cm01h01.w
C 105	20.2	80.8	620	3	BI404655	BI404655	MI-P-NA-a
C 106	20.2	80.8	633	9	AZ818427	AZ818427	2M0088N20
C 107	20.2	80.8	638	5	BY723544	BY723544	BY723544
C 108	20.2	80.8	644	9	CE071494	CE071494	t1gr-g88-
C 109	20.2	80.8	645	9	BH001660	BH001660	BMBAC01G1
C 110	20.2	80.8	650	3	BU661255	BU661255	BU661255
C 111	20.2	80.8	652	9	BZ459364	BZ459364	BONSH8TR
C 112	20.2	80.8	652	9	CE031150	CE031150	t1gr-g88-
C 113	20.2	80.8	675	9	BH001590	BH001590	BMBAC01E2
C 114	20.2	80.8	689	8	CX316160	CX316160	UG1.XZT66
C 115	20.2	80.8	699	9	BH939017	BH939017	oddb1c02.
C 116	20.2	80.8	701	9	AQ386471	AQ386471	RPC111-15
C 117	20.2	80.8	703	9	AZ989040	AZ989040	2M0272E16
C 118	20.2	80.8	708	10	BX173061	BX173061	Danlo rex
C 119	20.2	80.8	715	9	BH601721	BH601721	BOHMH47TF
C 120	20.2	80.8	730	3	BP119200	BP119200	BP119200
C 121	20.2	80.8	730	8	CX578594	CX578594	TTE000219
C 122	20.2	80.8	730	9	BZ912603	BZ912603	CH240.111
C 123	20.2	80.8	754	10	AG053182	AG053182	Pan treg1
C 124	20.2	80.8	756	6	CX316159	CX316159	UG1.XZT66
C 125	20.2	80.8	759	6	CA763874	CA763874	APF5-RPF
C 126	20.2	80.8	768	8	DN983298	DN983298	SV68F07.d
C 127	20.2	80.8	774	7	CF933586	CF933586	TTEST-B37
C 128	20.2	80.8	779	10	DU030218	DU030218	11326.Tom
C 129	20.2	80.8	780	9	BZ124836	BZ124836	CH230-439
C 130	20.2	80.8	786	9	BZ772088	BZ772088	mc85a08.
C 131	20.2	80.8	791	9	CC484028	CC484028	CH240.313
C 132	20.2	80.8	812	10	CL135873	CL135873	ISB1-107C
C 133	20.2	80.8	820	10	CL067511	CL067511	CH216-111
C 134	20.2	80.8	823	8	CX389935	CX389935	UG1.XZT37
C 135	20.2	80.8	827	8	CX389936	CX389936	UG1.XZT37
C 136	20.2	80.8	829	10	CZ209465	CZ209465	AlfA-aag8
C 137	20.2	80.8	832	9	BZ221007	BZ221007	CH230-372
C 138	20.2	80.8	834	9	CC332266	CC332266	OCTAM60TV
C 139	20.2	80.8	841	3	BP171067	BP171067	BP171067
C 140	20.2	80.8	843	10	BX968009	BX968009	Forward s
C 141	20.2	80.8	844	10	CG251504	CG251504	OGWGL45TH
C 142	20.2	80.8	854	9	CC675504	CC675504	OGAR57TV
C 143	20.2	80.8	862	9	CC536177	CC536177	CH240.415
C 144	20.2	80.8	865	11	CR094615	CR094615	Forward s
C 145	20.2	80.8	880	9	CC865385	CC865385	NDL.2P12.
C 146	20.2	80.8	899	10	CZ953237	CZ953237	265066.T0
C 147	20.2	80.8	931	10	CG251515	CG251515	OGWGL45TV
C 148	20.2	80.8	933	10	CL136450	CL136450	ISB1-108A
C 149	20.2	80.8	951	9	AQ747293	AQ747293	HS.5537.A
C 150	20.2	80.8	981	11	CNS03HGF	AL244248	Tetraodon

ALIGNMENTS

RESULT 1
CNS24164/c
LOCUS
DEFINITION
CNS24164
GQ015M11.T3 A06 GQ015 Populus trichocarpa x Populus deltoides cDNA
clone GQ015M11_A06 5', mRNA sequence.
ACCESSION
CNS24164
VERSION
CNS24164.1 GI:46842533
KEYWORDS
EST.
ORGANISM
Populus trichocarpa x Populus deltoides
Populus trichocarpa x Populus deltoides
Bukariya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.

REFERENCE
AUTHORS
1 (bases 1 to 202)
Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
Retzel,E. and Mackay,J.
TITLE
Arborea EST sequencing in Populus sp. (poplar)
JOURNAL
Unpublished (2004)
COMMENT
Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@ars.ullaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN Id Identifier: MN521358 Clone ID:
GQ015M11_A06 Clones available through: John Mackay, Ph. D.
professeur adjoint -Assistant professeur EMAIL:
jmackay@ars.ullaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: M11 row: 06 column: A
Seq primer: T3 primer.
Location/Qualifiers
1..202

FEATURES
source
/organism="Populus trichocarpa x Populus deltoides"
/mol_type="mRNA"
/strain="H1-11"
/db_xref="taxon:3695"
/clone="GQ015M11_A06"
/sex="MALE"
/tissue_type="Shoot tips and secondary stems of trees
approximately 80 cm tall."
/lab_host="E. coli DH10B cells"
/clone_lib="GQ015"
/note="Organ: Shoot tips and secondary stems of trees
approximately 80 cm tall; Vector: pBluescript II SK (+)
XR; Site 1: Eco-RI; Site 2: Xho-I; Shoot tips (including
apex, primary stem and developing leaves up to and
including LPI 1) and whole stems (wood and bark)
undergoing secondary growth taken from between LPI 8 and
LPI 15. Pooled sample of tissues harvested every two hours
during one 12 hour light/12 hour dark cycle, beginning 1
hour before the beginning of the light cycle, and ending 3
hours after the end of the light cycle. cDNA was prepared
from 5 Ng of poly A+ selected RNA and was directionally
ligated into the pluescript II SK (+) XR vector
(stratagene), transformed by electroporation into DH10B
cells (in vitro) for propagation"

ORIGIN

Query Match 87.2%; Score 21.8; DB 7; Length 202;
Best Local Similarity 92.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAAGCTGATGCTT 25
Db 199 AAAAAAAAAACGAAAGCTTGACTT 175

RESULT 2
AZ910928/c
LOCUS
DEFINITION
RPCI-24-114F16.TUB RPCI-24 Mus musculus genomic clone
AZ910928
ACCESSION
AZ910928
VERSION
AZ910928.1 GI:13229873
KEYWORDS
GSS.
ORGANISM
Mus musculus (house mouse)
Mus musculus
Bukariya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 282)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levine,M.,
Teagaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,

TITLE Russell, D., de Jong, P. and Fraser, C.M.
JOURNAL Mouse BAC End Sequences from Library RPCI-24
COMMENT Unpublished (1999)
Other_GSSs: RPCI-24-114F16.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 114 row: F column: 16
Seq primer: SP6
Class: BAC ends.

FEATURES

source Location/Qualifiers

1..282
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-114F16"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_id="RPCI-24"
/note="Vector: PTABAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pTABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 282;
Best Local Similarity 92.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTGATCTT 25
41 AAAAAAAAAAGCAAGCTTGATCTT 17

RESULT 3
LOCUS CW797493 321 bp DNA linear GSS 23-NOV-2004
DEFINITION WiscDslx0413-416P2 Arabidopsis thaliana T-DNA insertion flanking sequences Arabidopsis thaliana genomic, genomic survey sequence.
ACCESSION CW797493
VERSION CW797493.1 GI:55995321
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 321)
Woody, S.T., Monson, S.S., Jeeter, P.J., Austin-Phillips, S., Amasino, R.M., Susseman, M.R. and Krysan, P.J.
A New Community Resource for Knocking-out Small Genes and Tandemly-Duplicated Gene Families and for Mosaic Analysis in Arabidopsis
Unpublished (2004)
Contact: Woody ST
Biotechnology Center
University of Wisconsin-Madison
425 Henry Mall, Madison, WI 53706, USA
Tel: (608) 262-4640
Email: swoody@facstaff.wisc.edu
Class: TAIL-PCR.
Location/Qualifiers

source

1..321
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Col-0 ecotype"
/db_xref="taxon:3702"
/tissue_type="seeds produced by primary (Basta-resistant) transformants"
/clone_id="Arabidopsis thaliana T-DNA insertion flanking sequences"
/note="Vector: pDS-lox; Sequence generated in the course of an Arabidopsis T-DNA tagging program. TAIL-PCR was used to generate sequencing templates that represent A.T. genomic DNA flanking the left border of the pDS-lox T-DNA insert. PCR products were sequenced directly by using the p745 primer 5' AACGCCCAATGCTTTATTAAGTTGTC 3'."

ORIGIN

Query Match 85.6%; Score 21.4; DB 10; Length 321;
Best Local Similarity 95.7%; Pred. No. 2.2e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAGCTTGATCTT 25
Db 238 AAAAAAAAAAGCAAGCTTGATCTT 260

RESULT 4
LOCUS AV801140 394 bp mRNA linear EST 29-MAR-2002
DEFINITION AV801140 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-27-A16 3', mRNA sequence.
ACCESSION AV801140
VERSION AV801140.1 GI:19835125
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM
REFERENCE
AUTHORS Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 394)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rcc.riken.go.jp
An arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda PhiC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.
Location/Qualifiers
1..394
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-27-A16"
/dev_stage="Plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/clone_id="RAFL9"
/note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"

FEATURES

source

ORIGIN

Query Match 85.6%; Score 21.4; DB 1; Length 394;
 Best Local Similarity 95.7%; Pred. No. 2.2e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAGCTTAAGCTTGATCTT 25
 |||||
 2 AAAAAAGACTTAAGCTTGATCTT 24

RESULT 5
 BP603671 402 bp mRNA linear EST 25-JUN-2004
 LOCUS BP603671 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-59-J21 3',
 DEFINITION mRNA sequence.
 ACCESSION BP603671 GI:49233915
 VERSION BP603671
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE
 AUTHORS Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T., Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M., Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
 Functional annotation of a full-length Arabidopsis cDNA collection 1 (bases 1 to 402)
 1 (bases 1 to 402)
 Science 296 (5565), 141-145 (2002)

TITLE JOURNAL
 PUBMED 11910074
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msek@rtc.riken.go.jp
 reversed clone; please visit our web site
 (http://pfjweb.gsc.riken.go.jp/) for further details.

FEATURES
 source
 1. 402
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL16-59-J21"
 /lab_host="DH10B"
 /clone_lib="RAFL16"
 /note="Site_1: BamHI; Site_2: SalI; dark-grown"

ORIGIN

Query Match 85.6%; Score 21.4; DB 3; Length 402;
 Best Local Similarity 95.7%; Pred. No. 2.2e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAGCTTAAGCTTGATCTT 25
 |||||
 23 AAAAAAGACTTAAGCTTGATCTT 45

RESULT 6
 BP608064 419 bp mRNA linear EST 25-JUN-2004
 LOCUS BP608064 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-77-D23 3',
 DEFINITION mRNA sequence.
 ACCESSION BP608064 GI:49255579
 VERSION BP608064
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE
 AUTHORS Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T., Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M., Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
 Functional annotation of a full-length Arabidopsis cDNA collection 1 (bases 1 to 429)
 1 (bases 1 to 429)
 Science 296 (5565), 141-145 (2002)

TITLE JOURNAL
 PUBMED 11910074
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msek@rtc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda phage-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES
 Location/Qualifiers

REFERENCE
 AUTHORS Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T., Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M., Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
 Functional annotation of a full-length Arabidopsis cDNA collection 1 (bases 1 to 419)
 1 (bases 1 to 419)
 Science 296 (5565), 141-145 (2002)

TITLE JOURNAL
 PUBMED 11910074
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msek@rtc.riken.go.jp
 reversed clone; please visit our web site
 (http://pfjweb.gsc.riken.go.jp/) for further details.

FEATURES
 source
 1. 419
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL16-77-D23"
 /lab_host="DH10B"
 /clone_lib="RAFL16"
 /note="Site_1: BamHI; Site_2: SalI; dark-grown"

ORIGIN

Query Match 85.6%; Score 21.4; DB 3; Length 419;
 Best Local Similarity 95.7%; Pred. No. 2.2e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAGCTTAAGCTTGATCTT 25
 |||||
 23 AAAAAAGACTTAAGCTTGATCTT 45

RESULT 7
 AV795931 429 bp mRNA linear EST 29-MAR-2002
 LOCUS AV795931 RAFL8 Arabidopsis thaliana cDNA clone RAFL08-19-N06 3',
 DEFINITION mRNA sequence.
 ACCESSION AV795931 GI:19829914
 VERSION AV795931
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE
 AUTHORS Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
 Large scale analysis of Arabidopsis full-length cDNA (2002b)
 Unpublished (2002)
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msek@rtc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda phage-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified Bluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES
 Location/Qualifiers

```

source
1. 429
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL08-19-N06"
/dev_stage="rossette plants"
/lab_host="DH10B"
/clone_1lb="RAFL08"
/notes="Site_1: BamHI, Site_2: SalI; subjected to
dehydration-treated (1, 2, 5, 10, 24 hr)"

ORIGIN
Query Match      85.6%; Score 21.4; DB 1; Length 429;
Best Local Similarity 95.7%; Pred. No. 2.2e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAGTAAAGCTTGATCTT 25
    |||||
Db 152 AAAAAAAGTAAAGCTTGATCTT 130

RESULT 8
BP610473      447 bp      mRNA      linear      EST 26-JUN-2004
DEFINITION   BP610473 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-11-120 3',
LOCUS        BP610473
ACCESSION   BP610473
VERSION     BP610473.1
KEYWORDS    GI:49261655
SOURCE      Arabidopsis thaliana (chale cress)
ORGANISM    Arabidopsis thaliana
REFERENCE    Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 447)
Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M.,
Hayashida, T., Kawai, J., Carninci, P., Itoh, M., Ishi, Y.,
Arakawa, T., Shibata, K., Shingawa, A. and Shinzaki, K.,
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
11910074
PUBMED
COMMENT      Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: meeki@rc.riken.go.jp
reversed clone; please visit our web site
(http://pfweb.98c.riken.go.jp/) for further details.
Location/Qualifiers
source
1. 447
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL16-11-120"
/lab_host="DH10B"
/clone_1lb="RAFL16"
/notes="Site_1: BamHI, Site_2: SalI; dark-grown"

ORIGIN
Query Match      85.6%; Score 21.4; DB 3; Length 447;
Best Local Similarity 95.7%; Pred. No. 2.2e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAGTAAAGCTTGATCTT 25
    |||||
Db 23 AAAAAAAGTAAAGCTTGATCTT 45

RESULT 9
CX066556

```

```

LOCUS        CX066556      250 bp      mRNA      linear      EST 03-JAN-2005
DEFINITION   1321175 NCCCMW 04RT Oncorhynchus mykiss cDNA, mRNA sequence.
ACCESSION   CX066556
VERSION     CX066556.1
KEYWORDS    GI:56988122
SOURCE      Oncorhynchus mykiss (rainbow trout)
ORGANISM    Oncorhynchus mykiss
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 250)
Rexroad, C.E., Rise, M., Koop, B., von Schalburg, K. and Yao, J.,
04RT grool, NCCCMW/MVU EST Project, Phase II, in collaboration with
GRASP
Unpublished (2004)
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@ncccmw.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.
Plate: 104 row: L column: 21
Seq primer: ATTTAGGTGACACTATG.
Location/Qualifiers
1. 250
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/lab_host="top10"
/clone_1lb="NCCCMW 04RT"
/notes="Vector: PCR 4-TOPO. This is an early neurogenesis
SSH library created by Mathew L. Rise constructed by
subtracting late neurogenesis (mixed stages: hindbrain
swelling + heart tube with pericardialis) from early
neurogenesis (mixed stages: neural groove + 1/2 epiboly).
Fish were from a domesticated strain (Spring Valley Trout
Farm, Langley, B.C.), courtesy of Bob Devlin, DFO. These
are mostly internal (coding) sequences."

ORIGIN
Query Match      83.2%; Score 20.8; DB 8; Length 250;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAGTAAAGCTTGATCT 24
    |||||
Db 218 AAAAAAAGTAAAGCTTGATCT 241

RESULT 10
BX614017/c      251 bp      mRNA      linear      EST 07-AUG-2003
LOCUS        BX614017
DEFINITION   BX614017 Normalized Anopheles Head (NH) Library Anopheles gambiae
cDNA clone AGACD38TRB, mRNA sequence.
ACCESSION   BX614017
VERSION     BX614017.1
KEYWORDS    GI:33503904
SOURCE      EST.
ORGANISM    Anopheles gambiae (African malaria mosquito)
REFERENCE    Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Culicidae; Anophelinae; Anopheles.
1 (bases 1 to 251)
Lobo, N.L., Gardner, M., Romans, P. and Collins, F.H.,
Anopheles gambiae EST, Center for Tropical Disease Research and
Training
Unpublished (2003)
Contact: Frank H. Collins
Center for Tropical Disease Research and Training
University of Notre Dame

```

Notre Dame, IN 46556, USA

Tel: 574-631-9245

Fax: 574-631-3996

Email: frank.h.collins.75@nd.edu.

FEATURES

source

1. .251
Location/Qualifiers

/organism="Anopheles gambiae"

/mol_type="RNA"

/db_xref="taxon:7165"

/clone="AGACD387RB"

/lab_host="E. coli DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: EcoRI (5'end); Site_2: NotI (3'end); a

directionally cloned and normalized, oligo-T primed cDNA

library constructed from strain 4arr adult mosquito heads.

Equal numbers of sugar fed males, sugar fed females and 6,

24 and 48 hr post blood meal females were used: Bonaldo,

Lennon & Soares (1996): Normalization and Subtraction: Two

Approaches to Facilitate Gene Discovery, Genome Research

6, 791-806. ESTs sequenced from the M13 reverse priming

site reading from the 5' ends of the cDNAs are indicated

by 'R' in the clone name. ESTs sequenced from the M13

forward priming site reading from the 3' ends of the cDNAs

are indicated by 'F' in the clone name."

ORIGIN

Query Match

83.2%; Score 20.8; DB 5; Length 251;

Best Local Similarity 91.7%; Pred. No. 3.6e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24

Db 166 AAAAAAAAACTAAGCTTGATCT 143

RESULT 11

CR499737/c

LOCUS

CR499737 mch2-176N23RM1 BAC end, cultivar JamaLong A17 of Medicago

truncatula, genomic survey sequence.

ACCESSION CR499737

VERSION CR499737.1 GI:48661313

KEYWORDS

GSS. Medicago truncatula (barrel medic)

SOURCE

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;

Medicago.

1 (bases 1 to 271)

REFERENCE

AUTHORS

JOURNAL

Submitted (10-JUN-2004) Genoscope - Centre National de Sequencage;

BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr

- Web: www.genoscope.cns.fr)

Location/Qualifiers

1. .271

/organism="Medicago truncatula"

/mol_type="genomic DNA"

/cultivar="JamaLong A17"

/db_xref="taxon:3880"

/clone_lib="MTH2"

/note="Vector: pBelOBAC11; site_1: HindIII; site_2:

HindIII; Cook, D.R. and Kim, D.J

Genoscope sequence ID: mch2-176N23RM1"

ORIGIN

Query Match

83.2%; Score 20.8; DB 11; Length 271;

Best Local Similarity 91.7%; Pred. No. 3.6e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24

Db 205 AAAAAAAAACTAAGCTTGATCT 182

RESULT 12

CN292167

LOCUS

CN292167 409 bp mRNA linear EST 16-MAY-2004

17000583090086 GRN_PRENED Homo sapiens cDNA 5', mRNA sequence.

ACCESSION CN292167

VERSION CN292167.1 GI:47308581

KEYWORDS

EST.

SOURCE

Homo sapiens

Homo sapiens (human)

REFERENCE

AUTHORS

1 (bases 1 to 409)

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Flek, G., J.,

Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,

Lebkowski, J. and Stanton, L. W.

Transcriptome characterization elucidates signaling networks that

control human ES cell growth and differentiation

Nat. Biotechnol. 22 (6), 707-716 (2004)

15146197

COMMENT

Contact: Brandenberger R

Regenerative Medicine

Geron Corporation

230 Constitution Drive, Menlo Park, CA 94025, USA

Tel: 650 473 8658

Fax: 650 473 7760

Email: rbrandenberger@geron.com

Insert Length: 409 Std Error: 0.00.

Location/Qualifiers

1. .409

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/issue_type="embryonic stem cell, retinoic acid and

mitogen-treated hES cell line H7"

/clone_lib="GRN_PRENED"

/note="oligo dT primed, full-length enriched cDNA library

from hES cell line H7 (p29) maintained in feeder-free

conditions. Embryoid bodies were generated in the presence

of all-trans retinoic acid and mitogens."

ORIGIN

Query Match

83.2%; Score 20.8; DB 7; Length 409;

Best Local Similarity 91.7%; Pred. No. 3.6e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24

Db 153 AAAAAAAAACTAAGCTTGATCT 176

RESULT 13

CD451095/c

LOCUS

CD451095 479 bp mRNA linear EST 03-JUN-2003

USDA-FP 103150 Adult Alate Brown Citrus Aphid Toxoptera citricida

cDNA clone MHWTC-42_B06 5', mRNA sequence.

ACCESSION CD451095

VERSION CD451095.1 GI:31365835

KEYWORDS

EST.

Toxoptera citricida (brown citrus aphid)

Toxoptera citricida

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;

Aphidoidea; Aphididae; Aphidini; Toxoptera.

1 (bases 1 to 479)

Hunter, M.B., Dang, P.M., Baueher, M.G., Chaparro, J.X., McKendree, W.,

Shattner, R.G., Jr., McKenzie, C.L. and Sinsheimer, X.H.

Aphid biology: Expressed genes from alate Toxoptera citricida, the

brown citrus aphid

JOURNAL
COMMENT

J. Insect Sci. 3 (23), 1-7 (2003)
Contact: Wayne B. Hunter
US Horticultural Research Laboratory
USDA - ARS
2001 South Rock Rd., Fort Pierce, FL 34945, USA
Tel: (772) 462-5898
Fax: (772) 462-5960
Email: whunter@ehrl.ars.usda.gov
Seq primer: T3 primer.

FEATURES
source

Location/Qualifiers
1. .479
/organism="Toxoptera citricida"
/mol_type="mRNA"
/db_xref="taxon:223852"
/clone="HWTC-42_B06"
/sex="Mixed population"
/tissue_type="Entire insect"
/dev_stage="Adult Alate"
/lab_host="XLI-Blue"
/clone_lib="Adult Alate Brown Citrus Aphid"
/note="Vector: pBluescript II SK⁺; Site 1: EcoRI; Site 2: XhoI; Toxoptera citricida (Kirkaldy); A high quality EST with at least 100 contiguous bases at Trace Tuner score of 20 or better."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 479;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
|||||

Db 218 AAAAAAAAAACAAATGCTTGATCT 195
|||||

RESULT 14
BE157233/c 509 bp mRNA linear EST 21-JUN-2000
LOCUS RC4-HT0373-130200-011-c12 HT0373 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE157233
VERSION BE157233.1 GI:8620058
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 509)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zaigo, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

TITLE
JOURNAL
PUBMED
COMMENT

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4-HT0373-130200-011-c12&tl=2000-02-13&cl=1)
Seq primer: puc 18 forward
High quality sequence stop: 192.
Location/Qualifiers

FEATURES

source

1. .509
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0373"
/note="Organ: head, neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 509;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
|||||

Db 182 AAAAAAAAAAAAGCTTGATCT 159
|||||

RESULT 15
CC037473 528 bp DNA linear GSS 01-APR-2003
LOCUS 3591.1 87.1 H03.Y.1 3591 - Rescuemu Grid P Zea mays genomic.
DEFINITION genomic survey sequence.
ACCESSION CC037473
VERSION CC037473.1 GI:29452364
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade: Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 528)
Walbot, V.
Maize genomic sequences found using engineered Rescuemu transposon unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 3591.1 87.1 row: 22
Class: transposon-tagged.
Location/Qualifiers

FEATURES
source

1. .528
/organism="Zea mays"
/mol_type="genomic DNA"
/cultiivar="mixed background W23/A188/B73/K55"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="J3591 - Rescuemu Grid P"
/note="Organ: leaf; Vector: Rescuemu (engineered from pBluescript backbone); Site 1: BamHI; Site 2: BglII; Rescuemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescuemu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'Rescuemu.' Grid P was grown at Molokai in 2002. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 528;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 64 AAAAAAAAAAGTAATCTTGATCT 87
 |||||

Db 460 AAAAAAAAACAATGCTTGATCT 483
 |||||

RESULT 16
 CD450350 570 bp mRNA linear EST 03-JUN-2003
 LOCUS USDA-FP_102312 Adult Alate Brown Citrus Aphid Toxoptera citricida
 DEFINITION CDNA clone WHWTC-31_H08 5', mRNA sequence.
 ACCESSION CD450350
 VERSION CD450350.1 GI:31365090
 KEYWORDS EST.
 SOURCE Toxoptera citricida (brown citrus aphid)
 ORGANISM Toxoptera citricida (brown citrus aphid)
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
 Aphidoidea; Aphididae; Aphidini; Toxoptera.
 1 (bases 1 to 570)
 Hunter, W.B., Dang, P.M., Baueher, M.G., Chaparro, J.X., McKendree, W.,
 Shatters, R.G., Jr., McKenzie, C.L. and Sinister, X.H.
 Aphid biology: Expressed genes from alate Toxoptera citricida, the
 brown citrus aphid
 J. Insect Sci. 3 (23), 1-7 (2003)
 Contact: Wayne B. Hunter
 US Horticultural Research Laboratory
 USDA - ARS
 2001 South Rock Rd., Fort Pierce, FL 34945, USA
 Tel: (772) 462-5898
 Fax: (772) 462-5960
 Email: whunter@ushr1.ars.usda.gov
 Seq primer: T3 Primer.

FEATURES
 source
 1..570
 Location/Qualifiers
 /organism="Toxoptera citricida"
 /mol_type="mRNA"
 /db_xref="taxon:223852"
 /clone="WHWTC-31_H08"
 /sex="Mixed population"
 /tissue_type="Entire Insect"
 /dev_stage="Adult Alate"
 /lab_host="XJ1-Blue"
 /clone_1lb="Adult Alate Brown Citrus Aphid"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
 XhoI; Toxoptera citricida (Kirkaldy); A high quality EST
 with at least 100 contiguous bases at Trace Tuner score of
 20 or better."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 570;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 460 AAAAAAAAACAATGCTTGATCT 483
 |||||

RESULT 17
 DE083880 598 bp DNA linear GSS 25-MAY-2005
 LOCUS Oryzias latipes DNA, clone: ola1-185F06.R, genomic survey sequence.
 DEFINITION DE083880
 ACCESSION DE083880
 VERSION DE083880.1 GI:66717485
 KEYWORDS GSS.
 SOURCE Oryzias latipes (Japanese medaka)
 Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 AUTHORS Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.
 JOURNAL BAC end sequences of Olal Oryzias latipes library
 REFERENCE Published Only in Database (2005)
 2 (bases 1 to 598)
 Fujiyama, A.
 Direct Submission
 Submitted (20-MAY-2005) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
 (E-mail: afujiyama@gsc.riken.jp, URL: http://scg.gsc.riken.jp/,
 Tel: 81-3-4212-2558, Fax: 81-3-3556-1916)
 This work was done in collaboration with Takeda, H. (1), Naruse, K.
 (2)
 and Narita, T. (3)
 (1) Department of Biological Science,
 University of Tokyo
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
 Phone: +81-3-5841-4431
 Fax: +81-3-5841-4993
 E-mail: htakeeda.s.u-tokyo.ac.jp
 (2) Department of Biological Science,
 University of Tokyo
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
 Phone: +81-3-5841-4431
 Fax: +81-3-5841-4993
 E-mail: naruse.s.u-tokyo.ac.jp
 (3) Department of Biological Science,
 University of Tokyo
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
 Phone: +81-3-5841-4431
 Fax: +81-3-5841-4993
 E-mail: tanarita.s.u-tokyo.ac.jp

PRIMERS
 Sequencing : Reverse
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 L.Site 2 : SacI.
 Location/Qualifiers
 1..598
 /organism="Oryzias latipes"
 /mol_type="genomic DNA"
 /db_xref="taxon:8090"
 /clone="ola1-185F06.R"
 /sex="male"
 /cell_type="whole body"
 /clone_1lb="BAC end sequences of Olal Oryzias latipes
 library"

ORIGIN

Query Match 83.2%; Score 20.8; DB 11; Length 598;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 Db 505 AAAAAAAAACTAAGCTTTATTT 528
 |||||

RESULT 18
 CR535112 604 bp mRNA linear EST 07-JUL-2004
 LOCUS CR535112 Normalized Anopheles Head (NH) Library Anopheles gambiae
 DEFINITION CR535112
 ACCESSION CR535112
 VERSION CR535112.1 GI:49921276
 KEYWORDS EST.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preyigota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Anophelinae; Anopheles.

1 (bases 1 to 604)

Lobo, N.L., Gardner, M., Romans, P. and Collins, F.H.
Anopheles gambiae EST, Center for Tropical Disease Research and Training

JOURNAL
COMMENT
Unpublished (2003)
Contact: Frank H. Collins
Center for Tropical Disease Research and Training
University of Notre Dame
Notre Dame, IN 46556, USA
Tel: 574-631-9245
Fax: 574-631-3996
Email: Frank.h.collins.75@nd.edu
Contact: Frank H. Collins
Center for Tropical Disease Research and Training
University of Notre Dame, Notre Dame, IN 46556, USA. Tel: 574-631-9245
Fax: 574-631-3996
Email: Frank.h.collins.75@nd.edu.

FEATURES
source
1. .604
/organism="Anopheles gambiae"
/mol_type="mRNA"
/db_xref="taxon:7165"
/lab_host="E. coli DH10B"
/clone_id="Normalized Anopheles Head (NAH) Library"
/note="Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI (5' end); Site 2: NotI (3' end); a directionally cloned and normalized, oligo-T primed cDNA library constructed from strain 4air adult mosquito heads. Equal numbers of sugar fed males, sugar fed females and 6, 24 and 48 hr post blood meal females were used. Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery; Genome Research 6, 791-806. ESTs sequenced from the M13 reverse priming site reading from the 5' ends of the cDNAs are indicated by 'R' in the clone name. ESTs sequenced from the M13 forward priming site reading from the 3' ends of the cDNAs are indicated by 'F' in the clone name."

ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 604;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTATCT 24
166 AAAAAAAAACTAAGCTTATCT 143

Db

RESULT 19
BX953309 614 bp mRNA linear EST 01-MAR-2004
DKFZP781M0844_x1.781 (synonym: h1cc4) Homo sapiens cDNA clone
BX953309
BX953309 1 GI:43431901
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 614)
Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.
EST (Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R., et al.)
Unpublished (2003)

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZP781M0844) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14039 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1. .614
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZP781M0844"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="781 (synonym: h1cc4)"
/note="Vector: pSPORT1_Sfi; Site_1: SfiI; Site_2: SfiIb; cDNA-collection"

ORIGIN
Query Match 83.2%; Score 20.8; DB 5; Length 614;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTATCT 24
95 AAAAAAAAACTAAGCTTATCT 118

Db

RESULT 20
BU618293 632 bp mRNA linear EST 23-SEP-2002
BU618293
UI-H-FH1-bfh-o-19-0-UI.81 NCI CGAP FH1 Homo sapiens cDNA clone
UI-H-FH1-bfh-o-19-0-UI 3', mRNA sequence.
BU618293 1 GI:23284508
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 632)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA sequence: 47-80, >POLY A#Simple_repeat
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1. .632
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FH1-bfh-o-19-0-UI"
/tissue_type="Cell Line"
/dev_stage="Adult"

```

/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_FH1"
/note="Organ: Chondrosarcoma; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site_1: EcoR I;
Site_2: Not I; NCI CGAP FH1 is a normalized cDNA library
obtained from a cell line derived from grade I
chondrosarcoma tissue. The library was constructed and
normalized according to Bonaldi, Lennon and Soares, Genome
Research, 6:791-806, 1996. First strand cDNA synthesis was
primed with an oligo-dT primer containing a Not I site.
Double stranded cDNA was ligated to an EcoR I adaptor,
digested with Not I, and cloned directionally into
pT73-Pac vector. The oligonucleotide used to prime the
synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AGAAATCCGCGC. The cell line was provided by Dr. James Martin
from the University of Iowa.
TAG TISSUE=Human Chondrosarcoma Cell Line C58 - Grade 1
Chondrosarcoma
TAG_LIB=UI-H-FH1
TAG_SEQ=AGAAATCCGCGC"

ORIGIN

Query Match      83.2%; Score 20.8; DB 5; Length 632;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 66 AAAAAAAAAAAAAAAAAAGCTTGATCT 89

RESULT 21
CA429434      632 bp      mRNA      linear      EST 07-NOV-2002
LOCUS
DEFINITION
UI-H-FH1-bfm-d-01-0-UI.s1 NCI CGAP FH1 Homo sapiens cDNA clone
UI-H-FH1-bfm-d-01-0-UI 3', mRNA sequence.
ACCESSION
CA429434.1 GI:24792160
VERSION
CA429434.1
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catearrhini;
Homidae; Homo.
1 (bases 1 to 632)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: James Martin
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 47-80, >POLY AHSimple_repeat
Seq primer: M13 FORWARD
POLYA=yes.
Location/Qualifiers
1. 632
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FH1-bfm-d-01-0-UI"
/tissue_type="Cell Line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_FH1"

```

```

/note="Organ: Chondrosarcoma; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site_1: EcoR I;
Site_2: Not I; NCI CGAP FH1 is a normalized cDNA library
obtained from a cell line derived from grade I
chondrosarcoma tissue. The library was constructed and
normalized according to Bonaldi, Lennon and Soares, Genome
Research, 6:791-806, 1996. First strand cDNA synthesis was
primed with an oligo-dT primer containing a Not I site.
Double stranded cDNA was ligated to an EcoR I adaptor,
digested with Not I, and cloned directionally into
pT73-Pac vector. The oligonucleotide used to prime the
synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AGAAATCCGCGC. The cell line was provided by Dr. James Martin
from the University of Iowa.
TAG TISSUE=Human Chondrosarcoma Cell Line C58 - Grade 1
Chondrosarcoma
TAG_LIB=UI-H-FH1
TAG_SEQ=AGAAATCCGCGC"

ORIGIN

Query Match      83.2%; Score 20.8; DB 6; Length 632;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 66 AAAAAAAAAAAAAAAAAAGCTTGATCT 89

RESULT 22
CL358313/c      638 bp      DNA      linear      GSS 19-AUG-2004
LOCUS
DEFINITION
RPC144_414J7.f RPC1-44 Sus scrofa genomic clone RPC144_414J7,
genomic survey sequence.
ACCESSION
CL358313.1 GI:51410283
VERSION
CL358313.1
KEYWORDS
GSS.
SOURCE
Sus scrofa (pig)
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suidae;
Suidae;
1 (bases 1 to 638)
Sus.
1 (bases 1 to 638)
RPC144_414J7.f
Rogatcheva,M.B., Meyers,S., He,W., Larkin,D.M., Matron,B.M.,
Beever,J.E. and Schook,L.B.
PIGgy-BACing the Human Genome: Constructing a Porcine Physical Map
Through Comparative Genomics
Unpublished (2004)
Other_GSSs: RPC144_414J7.f
Contact: Lawrence B. Schook
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 265 5326
Fax: 217 244 5617
Email: schook@uiuc.edu
Clones are derived from the porcine BAC library RPC1-44
(http://www.bacpac.choi.org/porcine242.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@choi.org).
Clones may be purchased from BACPAC Resources
(http://BACPACorders.choi.org). This work was undertaken as part
of the International Swine Genome Sequencing Consortium by
University of Illinois at Urbana Champaign, USA with funds provided
by grant No. AG2002-34480-11828 from USDA-CSREES and
AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
Initiative)
Place: 414 row: J column: 7
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 638

```

FEATURES

source

/organism="Sus scrofa"
 /mol_type="genomic DNA"
 /strain="Four pigs (bred: 37.5% Yorks Landrace and 25%
 Meishan)"
 /db_xref="taxon:9823"
 /clone="RPC144_414J7"
 /sex="male"
 /cell_type="blood"
 /clone_id="RPC1-44"
 /note="Vector: pTARBAC2; site_1: EcoRI; site_2: EcoRI;
 porcine male BAC library produced by Pieter de Jong"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 638;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 Db 503 AAAAAAAAACTAAGCTTTTCT 480

RESULT 23 652 bp DNA linear GSS 01-FEB-2005
 AG920908
 LOCUS Drosophila auraria DNA, clone: DAB1-031F09.R.f.a, genomic survey
 DEFINITION
 ACCESSION AG920908
 VERSION AG920908.1 GI:58442279
 KEYWORDS GSS.
 SOURCE Drosophila auraria
 ORGANISM Drosophila auraria
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1
 Hattori, M., Toyoda, A., Murakami, K., Kuroki, Y., Fujiyama, A.,
 Toshio, T. K. and Sakaki, Y.
 BAC end sequences of library DAB1
 Unpublished
 2 (bases 1 to 652)
 Hattori, M.
 Direct Submission
 Submitted (19-JUN-2005) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
 (E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the BAC library DAB1
 For BAC library availability, please contact Masa-Toshi Yamamoto
 (yamamoto@kit.jp).
 Submitted (30-11-2004) by Masahira Hattori,
 RIKEN, Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: hattori@gsc.riken.jp, Tel: 81-45-503-9111,
 Fax: 81-45-503-9170)
 This work was done in collaboration with Yamamoto, M-T. Drosophila
 Genetic Resource Center
 Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8154, Japan
 Tel: 81-75-873-2660 FAX: 81-75-861-0881
 PRIMERS

COMMENT

Sequencing : R
 LIBRARY
 Vector : PKS150
 R.site 1 : SacI
 R.site 2 : SacI
 Location/Qualifiers
 1. 652
 /organism="Drosophila auraria"
 /mol_type="genomic DNA"
 /db_xref="taxon:47315"
 /clone="DAB1-031F09.R.f.a"
 /clone_lib="DAB1 Drosophila BAC library"

FEATURES

SOURCE

1. 652
 /organism="Drosophila auraria"
 /mol_type="genomic DNA"
 /db_xref="taxon:47315"
 /clone="DAB1-031F09.R.f.a"
 /clone_lib="DAB1 Drosophila BAC library"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 652;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTT 25
 |||||
 Db 428 AAAAAAAAACTAAGCTTGATTTT 451

RESULT 24

CE810279/c 659 bp DNA linear GSS 30-SEP-2003
 LOCUS tigr-gss-dog-17000317876511 Dog Library Canis familiaris genomic,
 DEFINITION
 ACCESSION CE810279
 VERSION CE810279.1 GI:37151266
 KEYWORDS GSS.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 Canis.

REFERENCE 1 (bases 1 to 659)
 Kirkness, E.F., Balda, V., Halpern, A.L., Levy, S., Remington, K.,
 Ruesch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
 Venter, J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 14512627
 Contact: Kirkness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirkness@tigr.org
 Class: shotgun.

FEATURES

source

1. 659
 Location/Qualifiers
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BstXI; Libraries were prepared from
 peripheral blood"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 659;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 Db 222 AAAAAAAAAAGAAAGCTTGATCT 199

RESULT 25 666 bp DNA linear GSS 25-SEP-2003
 CE201628/c
 LOCUS tigr-gss-dog-17000372234139 Dog Library Canis familiaris genomic,
 DEFINITION
 ACCESSION CE201628
 VERSION CE201628.1 GI:35357283
 KEYWORDS GSS.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 Canis.

REFERENCE 1 (bases 1 to 666)

AUTHORS
 Kirckness, E. F., Bfna, V., Halpern, A. L., Levy, S., Remington, K., Ruch, D. B., Delcher, A. L., Pop, M., Wang, W., Fraser, C. M. and Venter, J. C.
TITLE
 The dog genome: survey sequencing and comparative analysis
JOURNAL
 Science 301 (5641), 1898-1903 (2003)
PUBMED
 14512627
COMMENT
 Contact: Kirckness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirkness@tigr.org
 Class: shotgun.
FEATURES
 Location/Qualifiers
 source
 1..666
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="Standard Poodle"
 /db_xref="taxon:9615"
 /clone_lib="Dog Library"
 /note="Site 1: BstXI; Libraries were prepared from peripheral blood"
ORIGIN
 Query Match 83.2%; Score 20.8; DB 9; Length 666;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY
 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 153 AAAAAAAAAATAAGCTTGATCT 130
RESULT 26
 LOCUS CR629796 671 bp mRNA linear EST 11-AUG-2004
 DEFINITION DKFZp46961722.F1.469 (synonym: pkid1) Pongo pygmaeus cDNA clone
 ACCESSION CR629796
 VERSION CR629796
 KEYWORDS EST.
 SOURCE CR629796.1 GI:51125876
 ORGANISM Pongo pygmaeus (orangutan)
 Pongo pygmaeus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pongo.
 1 (bases 1 to 671)
 Bahr, A., Lauber, J., Mewes, H. W., Well, B., Amid, C., Osanger, A., Fodor, G., Han, M. and Wiemann, S.
 Pongo pygmaeus mRNA (Bahr, A., Lauber, J., Mewes, H. W., et al.)
 Unpublished (2004)
 Contact: MIPS
TITLE
 Ingolstaedter Landstr. 1, D-85764 Neuberg, Germany
 This is the 5' sequence of the clone insert from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp46961722) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering:
 http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneId=DKFZp46961722
 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
FEATURES
 Location/Qualifiers
 source
 1..671
 /organism="Pongo pygmaeus"
 /mol_type="mRNA"
 /db_xref="taxon:9600"
 /clone="DKFZp46961722"
 /tissue_type="kidney"
 /dev_stage="adult"
ORIGIN
 Query Match 83.2%; Score 20.8; DB 7; Length 671;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY
 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 94 AAAAAAAAAATAAGCTTGATCT 117
RESULT 27
 LOCUS AG047569 674 bp DNA linear GSS 02-NOV-2001
 DEFINITION Pan troglodytes DNA, clone: PTB-027C03.F, genomic survey sequence.
 ACCESSION AG047569
 VERSION AG047569.1 GI:16584461
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.
 1
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of library PTB
 Unpublished
 2 (bases 1 to 674)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenho-chou, Tsukuba-shi, Ibaraki, Japan (E-mail: chimpes@gsc.riken.go.jp, URL: http://ngp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.
PRIMERS
 Sequencing: -21M13
LIBRARY
 Vector : pKS145
 R Site 1 : SacI
 R Site 2 : SacI.
Location/Qualifiers
 1..674
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="PTB-027C03.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
 Query Match 83.2%; Score 20.8; DB 10; Length 674;
 Best Local Similarity 91.7%; Pred. No. 3.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY
 1 AAAAAAAAACTAAGCTTGATCT 24
 |||||
 529 AAAAAAAAAATAAGCTTGATTT 552
RESULT 28
 LOCUS AG051677 675 bp DNA linear GSS 02-NOV-2001
 DEFINITION Pan troglodytes DNA, clone: PTB-033M10.F, genomic survey sequence.

```

ACCESSION   AG051677
VERSION     AG051677.1  GI:16569120
KEYWORDS
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominoidea; Pan.
REFERENCE   1
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Tokoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      BAC end sequences of library PTB
JOURNAL     Unpublished
AUTHORS     2 (bases 1 to 675)
            Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Tokoki, Y., Watanabe, H. and Sakaki, Y.
REFERENCE   Direct Submission
JOURNAL     Submitted (02-AUG-2001) Ageo Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail: chimpbes@gsc.riken.go.jp, URL: http://hsp.gsc.riken.go.jp/,
            TEL: 81-45-503-9111, Fax: 81-45-503-9170)
            Clones are derived from the chimpanzee BAC library PTB. This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
COMMENT
PRIMERS
Sequencing: -21M13
LIBRARY
Vector      : pKS145
R.Site 1    : SacI
R.Site 2    : SacI
FEATURES
source
1. 675
    /organism="Pan troglodytes"
    /mol_type="genomic DNA"
    /db_xref="taxon:9598"
    /clone="PTB-033M10.F"
    /sex="male"
    /cell_type="lymphoblast"
    /clone_id="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match      83.2%; Score 20.8; DB 10; Length 675;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAAGCTTGATCT 24
    |||||
Db 528 AAAAAAAAACTAAAGCTTGATTT 551
RESULT 29
AL599901/c 683 bp mRNA linear EST 04-SEP-2003
LOCUS      DKFZp313p2331_r1 313 (synonym: nlcc2) Homo sapiens cDNA clone
DEFINITION DKFZp313p2331 5', mRNA sequence.
ACCESSION   AL599901
VERSION     AL599901.1 GI:15163189
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominoidea; Homo.
REFERENCE   1 (bases 1 to 683)
AUTHORS    Mambut, R., Heubner, D., Mewes, W., Weil, B. and Wiemann, S.
TITLE      EST (Mambut, R., Heubner, D., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL     Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

```

```

FEATURES
source
1. 683
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="DKFZp313p2331"
    /dev_stage="adult"
    /lab_host="DH10B"
    /clone_id="313 (synonym: nlcc2)"
    /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
            cDNA-collection"
ORIGIN
Query Match      83.2%; Score 20.8; DB 1; Length 683;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAAGCTTGATCT 24
    |||||
Db 675 AAAAAAAAACTAAAGCTTGATCT 652
RESULT 30
AL599901/c 709 bp DNA linear GSS 27-OCT-2004
LOCUS      CW631677
DEFINITION OP_Ba0058F09.f OP_Ba Oryza punctata genomic clone OP_Ba0058F09
ACCESSION   CW631677
VERSION     CW631677.1 GI:54665639
KEYWORDS    GSS.
SOURCE      Oryza punctata
ORGANISM    Eukaryota; Viridiplantae; Scrophophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
            1 (bases 1 to 709)
            SamMiguel, P., Westerman, R., Kim, H., Yu, Y., Wisotski, M., Yost, D.,
            Stum, D., Rao, K., Luo, M., Jetty, R., Kudrna, D., Miller, C.,
            Hatfield, J., Soderlund, C., Wing, R. and Jackson, S.A.
            OMAP Project - Purdue University
            Unpublished (2004)
            Contact: Scott A. Jackson
            Jackson Laboratory
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with Lucy version 1.19e.
            Bases 42-750 of the raw sequence (length 1079) were retained after
            clipping.
            PCR Primers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert length: 161 Std Error: 0.00
            Plate: 0058 row: F column: 09
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES
source
1. 709
    /organism="Oryza punctata"
    /mol_type="genomic DNA"
    /db_xref="taxon:4537"
    /clone="OP_Ba0058F09"
    /issue_type="young leaves"

```

ORIGIN

/lab_host="DH10B-T1 phage resistant"
/clone_lib="Op_Ba"
/note="Vector: pGIBAC1; Site_1: HindIII; Site_2: HindIII"

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 10; Length 709;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

2 AAAAAAAAACTAAAGCTTGATCT 25
389 AAAAAAAAACTAAAGCTTGTTT 412

DB

RESULT 31
LOCUS BM633829/c 712 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687508088 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
19600449665723 5', mRNA sequence.

ACCESSION

BM633829 GI:18933340

VERSION

BM633829.1 GI:18933340

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU0100417W row: E column: 17
Seq primer: M13 Reverse.
Location/Qualifiers

FEATURES

source

1..712
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449665723"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPori1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSPori 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mt4.org)."

ORIGIN

Query Match 83.2%; Score 20.8; DB 3; Length 712;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
537 AAAAAAAAACTAAATTGATCT 514

DB

RESULT 32
LOCUS CR769513 728 bp mRNA linear EST 23-SEP-2004
DEFINITION DKFZp469H2329_r1_469 (synonym: pkid1) Pongo pygmaeus cDNA clone
DKFZp469H2329_5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CR769513 GI:52613541
EST.
Pongo pygmaeus (orangutan)
Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Pongo.
1 (bases 1 to 728)
Ostenwelder, B., Obermaier, B., Deutschenbauer, S., Schaipp, A.,
Mewes, H.W., Weill, B., Amid, C., Oesinger, A., Fobo, G., Han, M. and
Wiemann, S.
Pongo pygmaeus mRNA (Ostenwelder, B., Obermaier, B.,
Deutschenbauer, S., et al.)
Unpublished (2004)
Contact: MIPS
MIPS

Ingolstaedter Landstr.1, D-85764 Neuberg, Germany
This is the 5' sequence of the clone insert. Clone from S. Wiemann,
Molecular Genome Analysis, German Cancer Research Center (DKFZ);
Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix
(Martinsried/Germany) within the cDNA sequencing consortium of the
German Genome Project. This clone (DKFZp469H2329) is available at
the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
Berlin, Germany. Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneid=DKFZp469H2329
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/.

FEATURES

source

1..728
/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp469H2329"
/tissue_type="Kidney"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="469 (synonym: pkid1)"
/note="Vector: pSPori1_Sfi; Site_1: SfiI; Site_2: SfiIb"

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 728;
Best Local Similarity 91.7%; Pred. No. 3.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
65 AAAAAAAAACTAAAGCTTGATCT 88

DB

RESULT 33
LOCUS CK404334/c 813 bp mRNA linear EST 05-JUN-2004
DEFINITION AUF IfdhK_245_b16 Ictalurus furcatus head kidney cDNA library
Ictalurus furcatus cDNA 5', mRNA sequence.
CK404334
1 GI:40563574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Ictalurus furcatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Ictaluridae; Ictalurus.
1 (bases 1 to 813)
Liu, Z., Li, P., Liu, L., He, C., Kuang, H., Feng, J., Chen, L.,
Peatman, E., Bao, P., Li, L., He, C., Kuang, H., Feng, J., Chen, L.,
Dunham, R. and Brady, Y.
30,000 new catfish ESTs: new resources for functional analysis of
genes involved in aquaculture performance traits
Unpublished (2004)
Contact: Liu ZJ
The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell

and Molecular Biosciences

Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA

Tel: 334 844 4054

Fax: 334 844 9208

Email: zliu@aceag.auburn.edu

Seq primer: T7.

FEATURES

source

Location/Qualifiers
1..813
/organism="Ictalurus furcatus"
/mol_type="mRNA"
/db_xref="taxon:66913"
/clone_lib="Ictalurus furcatus head kidney cDNA library"
/note="Organ: Head kidney; Vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 7; Length 813;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 AAAAAAAAACTAAAGCTGATCT 24
|||||
691 AAAAAAAAAAAAGCTGATCT 668

Db

RESULT 34

CG943200/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CG943200 848 bp DNA linear GSS 15-DEC-2003
MBENNS1TF mch2 Medicago truncatula genomic clone 916, genomic
survey sequence.

CG943200
CG943200.1 GI:39847813

GSS.
Medicago truncatula (barrel medic)

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

Rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifoliales;

Medicago.

1 (bases 1 to 848)

Town, C.D., Shetty, J., Koo, H. and Feldblyum, T. F.

Sequencing of BAC ends from Medicago truncatula

Unpublished (2003)

Other_GSSs: MBENNS1TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

Seq primer: TGTAAACGACGCGCAGT

Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..848
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone_lib="916"
/clone_lib="mch2"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII; Cook, D.R. and Kim, D.J. unpublished"

ORIGIN

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 10; Length 848;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 AAAAAAAAACTAAAGCTGATCT 24
|||||
583 AAAAAAAAACTAAAGCTGATCT 560

Db

RESULT 35

CL119775

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CL119775 920 bp DNA linear GSS 05-JAN-2004
ISB1-77023 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-77023,
genomic survey sequence.
CL119775
CL119775.1 GI:40613410
GSS.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGG
Class: BAC ends
High quality sequence start: 17
High quality sequence stop: 674.
Location/Qualifiers
1..920
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8164"
/clone="ISB1-77023"
/clone_lib="ISB1"
/note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

FEATURES

source

ORIGIN

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 10; Length 920;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 AAAAAAAAACTAAAGCTGATCT 24
|||||
573 AAAAAAAAACTAAAGCTGATCT 596

Db

RESULT 36

AQ239678/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AQ239678 472 bp DNA linear GSS 30-SEP-1998
CIT-HSP-2387N14.TF.1 CIT-HSP Homo sapiens genomic clone 2387N14,
genomic survey sequence.
AQ239678
AQ239678.1 GI:3671969
GSS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 472)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J. C.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1998)
Map Building
Other GSSs: CIT-HSP-2387N14.TF.1
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@ctigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.ctigr.org/cdb/hunguen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..472
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="2387N14"
 /sex="Male"
 /cell_type="Sperm"
 /clone_1ib="CIT-HSP"
 /note="Vector: pBelBAC11, Site_1: HindIII, Site_2:
 HindIII"

ORIGIN

Query Match 81.6%; Score 20.4; DB 9; Length 472;
 Best Local Similarity 95.5%; Pred. No. 5e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGAT 22
 |||||
 266 AAAAAAAAACTAAGCTTGAT 245

RESULT 37
 B2469233 573 bp DNA linear GSS 13-DEC-2002
 LOCUS B0NFJ26TR.B0.1.6.2 KB.tot Brassica oleracea genomic clone B0NFJ26,
 DEFINITION genomic survey sequence.
 ACCESSION B2469233
 VERSION B2469233.1 GI:26765021
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 573)
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
 Utecherback,T.R., Mortman,J.R., White,O.R. and Town,C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)

JOURNAL
 PUBMED 15805490
 COMMENT Other GSSs: B0NFJ26TF
 Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@ctigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES

source

Location/Qualifiers
 1..573
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO100DH3"
 /db_xref="taxon:3712"
 /clone="B0NFJ26"
 /clone_1ib="BO_1.6.2_KB.tot"
 /note="Vector: pHOS1, Site 1: BstXI, 1.6-2 kb sheared
 total DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match 81.6%; Score 20.4; DB 9; Length 573;

Best Local Similarity 95.5%; Pred. No. 5e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAAGCTTGAT 22
 |||||
 510 AAAAAAAAACTAAGCTTGAT 531

RESULT 38
 BH736678/c 580 bp DNA linear GSS 20-FEB-2002
 LOCUS B0M1A17R.B0.2.3_KB Brassica oleracea genomic clone B0M1A17,
 DEFINITION genomic survey sequence.
 ACCESSION BH736678
 VERSION BH736678.1 GI:18842073
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 580)
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
 Utecherback,T.R., Mortman,J.R., White,O.R. and Town,C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)

JOURNAL
 PUBMED 15805490
 COMMENT Other GSSs: B0M1A17TF
 Contact: Chris Town

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@ctigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES

source

Location/Qualifiers
 1..580
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO100DH3"
 /db_xref="taxon:3712"
 /clone="B0M1A17"
 /clone_1ib="BO_2.3_KB"
 /note="Vector: pHOS1, Site 1: BstXI, 2-3 kb sheared
 genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match 81.6%; Score 20.4; DB 9; Length 580;
 Best Local Similarity 95.5%; Pred. No. 5e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGAT 22
 |||||
 570 AAAAAAAAACTAAGCTTGAT 549

RESULT 39
 BH993161 715 bp DNA linear GSS 07-OCT-2002
 LOCUS ce161b10.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
 DEFINITION sequence.
 ACCESSION BH993161
 VERSION BH993161.1 GI:23534051
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 715)

REFERENCE

AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rabinowicz, P.D. and Wilson, R.K.
TITLE Whole genome shotgun reads from *Brassica oleracea*
JOURNAL Unpublished (2002)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.wustl.edu
Plate: oel61 row: b column: 10
Seq primer: -21UPPOT forward
Class: shotgun
High quality sequence start: 44
High quality sequence stop: 551.
Location/Qualifiers
1..715
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using *Brassica oleracea* T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN
Query Match 81.6%; Score 20.4; DB 9; Length 715;
Best Local Similarity 95.5%; Pred. No. 5e+03; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGAT 22
|||||
82 AAAAAAAAAATTAAAGCTTGAT 103
|||||

RESULT 40
BH602405 735 bp DNA linear GSS 15-DEC-2001
LOCUS B0G1B42TF B0G1 *Brassica oleracea* genomic clone B0G1B42, genomic
DEFINITION Survey sequence.
ACCESSION BH602405
KEYWORDS BH602405.1 GI:17854851
SOURCE GSS.
ORGANISM *Brassica oleracea*
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; *Brassica*.
1 (bases 1 to 735)
Ayele, M., Haas, B.U., Kumar, N., Wu, H., Xiao, Y., Van Aken, S.,
Utecherack, T.R., Mortman, J.R., White, O.R. and Town, C.D.
Whole genome shotgun sequencing of *Brassica oleracea* and its
application to gene discovery and annotation in *Arabidopsis*
Genome Res. 15 (4), 487-495 (2005)
JOURNAL PUBMED
PUBMED 15805490
COMMENT Other GSSes: B0G1B42TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..735
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"

FEATURES
Source

/clone="B0G1B42"
/clone_lib="B0G1"
/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
Query Match 81.6%; Score 20.4; DB 9; Length 735;
Best Local Similarity 95.5%; Pred. No. 5e+03; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGAT 22
|||||
568 AAAAAAAAAATTAAAGCTTGAT 589
|||||

RESULT 41
BZ025719/c 737 bp DNA linear GSS 08-OCT-2002
LOCUS oeh92a02.g1 *B.oleracea002* *Brassica oleracea* genomic, genomic survey
DEFINITION sequence.
ACCESSION BZ025719
VERSION BZ025719.1 GI:23586953
KEYWORDS GSS.
SOURCE *Brassica oleracea*
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; *Brassica*.
1 (bases 1 to 737)
Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rabinowicz, P.D. and Wilson, R.K.
TITLE Whole genome shotgun reads from *Brassica oleracea*
JOURNAL Unpublished (2002)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.wustl.edu
Plate: oeh92 row: a column: 02
Seq primer: -28RPOT reverse
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 551.
Location/Qualifiers
1..737
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using *Brassica oleracea* T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN
Query Match 81.6%; Score 20.4; DB 9; Length 737;
Best Local Similarity 95.5%; Pred. No. 5e+03; Indels 1; Gaps 0;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGAT 22
|||||
449 AAAAAAAAAATTAAAGCTTGAT 428
|||||

RESULT 42
CL651769/c 738 bp DNA linear GSS 09-JUL-2004
LOCUS PRI0113a C08 - PRI0113a.B21 (738) Mixed stage fosmid library of *P.*
DEFINITION *pacificus* var. *Californica* *Pristionchus pacificus* genomic, genomic
survey sequence.
ACCESSION CL651769

```

VERSION      CL651769.1  GI:50130396
KEYWORDS     GSS.
SOURCE       Pristionchus pacificus
ORGANISM     Pristionchus pacificus
              Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
              Neodiplogasteridae; Pristionchus.
REFERENCE    1 (bases 1 to 738)
              Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
              AppADB: an AceDB database for the nematode satellite organism
              Pristionchus pacificus
JOURNAL      Nucleic Acids Res. 32 (1), D421-D422 (2004)
PUBMED       14681447
COMMENT      Contact: Sommer RJ
              Evolutionary Biology
              Max-Planck-Institute for Developmental Biology
              Spemannstr. 37-39, Tuebingen D-72076, Germany
              Tel: 00497071601371
              Fax: 00497071601498
              Email: ralf.sommer@tuebingen.mpg.de
              This library was generated at Caltech, Pasadena, USA and end
              sequenced at Vancouver, Canada.
              Seq primer: T7
              Class: fosmid ends.
FEATURES     Location/Qualifiers
              1..738
              /organism="Pristionchus pacificus"
              /mol_type="genomic DNA"
              /strain="California"
              /db_xref="taxon:54126"
              /clone_lib="Mixed stage fosmid library of P. pacificus
              var. California"
              /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN
Query Match      81.6%; Score 20.4; DB 10; Length 738;
Best Local Similarity 95.5%; Pred. No. 5e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGAT 22
    |||||
    478 AAAAAAAAACTTAAGCTTGAT 457

RESULT 43
DR922978/c      741 bp  mRNA  linear  EST 02-AUG-2005
DEFINITION     EST114517 Aquilegia cDNA library Aquilegia formosa x Aquilegia
                pubescens cDNA clone COLMT32, mRNA sequence.
ACCESSION      DR922978
VERSION        DR922978.1  GI:71692341
KEYWORDS       EST.
SOURCE         Aquilegia formosa x Aquilegia pubescens
                Aquilegia formosa x Aquilegia pubescens
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
                Ranunculaceae; Aquilegia.
REFERENCE      1 (bases 1 to 741)
              Hodges,S.A., Rensink,W., Bueli,C.R., Borevitz,J., Kramer,E.,
              Nordborg,M. and Tomkins,J.
              Generation of ESTs from Aquilegia
              Unpublished (2005)
              Other_ESTs: EST114516
              Contact: Scott Hodges
              Department of Ecology, Evolution and Marine Biology
              University of California, Santa Barbara
              Santa Barbara, CA 93106, USA
              Tel: 805 893 7813
              Fax: 805 893 4724
              Email: hodges@lifesci.ucsb.edu
              Seq primer: M13 Reverse.
              Location/Qualifiers
              1..741
              /organism="Aquilegia formosa x Aquilegia pubescens"

```

```

/mol_type="mRNA"
/db_xref="taxon:338618"
/clone="COLMT32"
/tissue_type="mixed shoot and floral apical meristems,
flower buds, leaves and roots"
/lab_host="DH10B T1 (T1 and T5 phage resistance)"
/clone_lib="Aquilegia cDNA library"
/note="Vector: pCMV SPORT6.1; Site_1: EcoRI, Site_2: NotI;
F2, F3, and F4 lines of Aquilegia formosa x A. pubescens
were grown from seed in greenhouses at UC Santa Barbara.
From these plants three sets of tissue were collected: 1)
Small flower buds (<10 mm) and very young inflorescences
(71 & 29% by weight respectively), 2) Medium (7-20 mm) and
large (at or near anthesis) flower buds (65 & 35% by
weight respectively) and 3) Shoot apical meristems. A
fourth set of tissue was collected from plants of A.
formosa. These plants were grown from seed in sand and at
approximately 1 month root tissue and leaf tissue of
various developmental stages were collected (84 & 16% by
weight respectively). Total RNA was extracted from each
set of tissue and pooled in the following proportions:
1.5X from sets 1 & 2, 1X from sets 3 & 4. From the pooled
total RNA, mRNA was extracted and enriched for full-length
messages and then normalized with proprietary methods by
Invitrogen."

ORIGIN
Query Match      81.6%; Score 20.4; DB 8; Length 741;
Best Local Similarity 95.5%; Pred. No. 5e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGAT 22
    |||||
    143 AAAAAAAAAATTAAGCTTGAT 122

RESULT 44
BH598221/c      761 bp  DNA  linear  GSS 15-DEC-2001
LOCUS          BH598221
DEFINITION     BCGYH92TF BCGY Brassica oleracea genomic clone BCGYH92, genomic
                survey sequence.
ACCESSION      BH598221
VERSION        BH598221.1  GI:17850673
KEYWORDS       GSS.
SOURCE         Brassica oleracea
                Brassica oleracea
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
                rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE      1 (bases 1 to 761)
              Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
              Utecherback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
              Whole genome shotgun sequencing of Brassica oleracea and its
              application to gene discovery and annotation in Arabidopsis
              Genome Res. 15 (4), 487-495 (2005)
              15805490
              Other_GSSs: BCGYH92TF
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seq primer: TF
              Class: sheared ends.
              Location/Qualifiers
              1..761
              /organism="Brassica oleracea"
              /mol_type="genomic DNA"
              /strain="TO1000DH3"
              /db_xref="taxon:3712"
              /clone="BCGYH92"

```



```

/clone lib="BOGY"
/notes=Vector: PHOS1, Site 1: BexXI, 2-3 kb sheared
genomic DNA inserted into PHOS1 using BexXI linkers"
ORIGIN
Query Match      81.6%; Score 20.4; DB 9; Length 761;
Best Local Similarity 95.5%; Pred. No. 5e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGAT 22
   ||||| ||||| ||||| |||||
Db 685 AAAAAAAAAATAAGCTTGAT 664

RESULT 45
BZ017481/c 764 bp DNA linear GSS 08-OCT-2002
LOCUS oed93a11.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION
ACCESSION BZ017481.1 GI:23576047
VERSION BZ017481.1
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 764)
Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
Nash,W., Rabinowicz,P.D. and Wilson,R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
CONTACT: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: oed93 row: e column: 11
Seq primer: -28RPOT reverse
Class: shotgun
High quality sequence start: 87
High quality sequence stop: 479.
Location/Qualifiers
1..764
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/notes=Vector: POTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea TO1000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
ORIGIN
Query Match      81.6%; Score 20.4; DB 9; Length 764;
Best Local Similarity 95.5%; Pred. No. 5e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGAT 22
   ||||| ||||| ||||| |||||
Db 180 AAAAAAAAAATAAGCTTGAT 159

RESULT 46
CL671117/c 819 bp DNA linear GSS 09-JUL-2004
LOCUS PRI0163d.G04 - PRI0163d.B21 (819) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION CL671117
VERSION CL671117.1 GI:50169772.

```

```

KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 819)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447
PUBMED
JOURNAL
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
Location/Qualifiers
1..819
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/notes=Vector: pepifos-5 Fosmid vector"
ORIGIN
Query Match      81.6%; Score 20.4; DB 10; Length 819;
Best Local Similarity 95.5%; Pred. No. 5e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGAT 22
   ||||| ||||| ||||| |||||
Db 585 AAAAAAAAAAGCTTAAGCTTGAT 564

RESULT 47
DR951334/c 876 bp mRNA linear EST 02-AUG-2005
LOCUS EST1142873 Aguillegia cDNA library Aguillegia formosa x Aguillegia
DEFINITION pubescens cDNA clone COIRT92, mRNA sequence.
ACCESSION DR951334
VERSION DR951334.1 GI:71720697
KEYWORDS EST.
SOURCE Aguillegia formosa x Aguillegia pubescens
ORGANISM Aguillegia formosa x Aguillegia pubescens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;
Ranunculaceae; Aquilegia.
REFERENCE 1 (bases 1 to 876)
Hodges,S.A., Rensink,W., Buell,C.R., Borevitz,J., Kramer,E.,
Nordborg,M. and Tomkins,J.
Generation of ESTs from Aguillegia
Unpublished (2005)
Other_ESTs: EST1142872
Contact: Scott Hodges
Department of Ecology, Evolution and Marine Biology
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
Tel: 805 893 7813
Fax: 805 893 4724
Email: hodges@lifesci.ucsb.edu
Seq primer: M13 Reverse.
Location/Qualifiers
1..876
/organism="Aguillegia formosa x Aguillegia pubescens"
/mol_type="mRNA"
FEATURES
source

```

/db_xref="taxon:338618"
 /clone="COIR792"
 /tissue_type="mixed shoot and floral apical meristems,
 flower buds, leaves and roots"
 /lab_host="DH10B T1 (T1 and T5 phage resistance)"
 /clone_11b="Aquillegia cDNA library"
 /note="Vector: pCMV SPORTS.1; Site 1: EcoRI; Site 2: NotI;
 F2, F3, and F4 lines of Aquilegia formosa x A. pubescens
 were grown from seed in greenhouses at UC Santa Barbara.
 From these plants three sets of tissue were collected: 1)
 Small flower buds (<10 mm) and very young inflorescences
 (71 & 294 by weight respectively); 2) Medium (7-20 mm) and
 large (at or near anthesis) flower buds (65 & 354 by
 weight respectively) and 3) Shoot apical meristems. A
 fourth set of tissue was collected from plants of A.
 formosa. These plants were grown from seed in sand and at
 approximately 1 month root tissue and leaf tissue of
 various developmental stages were collected (84 & 164 by
 weight respectively). Total RNA was extracted from each
 set of tissue and pooled in the following proportions:
 1.5X from sets 1 & 2, 1X from sets 3 & 4. From the pooled
 total RNA, mRNA was extracted and enriched for full-length
 messages and then normalized with proprietary methods by
 Invitrogen."

ORIGIN

Query Match 81.6%; Score 20.4; DB 8; Length 876;
 Best Local Similarity 95.5%; Pred. No. 5e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGAT 22
 87 AAAAAAAAAATAAGCTTGAT 66

RESULT 48
 AJ856529/c 1338 bp DNA linear GSS 11-NOV-2004
 LOCUS AJ856529
 DEFINITION Brassica rapa subsp. pekinensis GSS, clone KBRH053P05, end read,
 primer M13(forward), genomic survey sequence.
 ACCESSION AJ856529.1 GI:55653852
 VERSION
 KEYWORDS GSS; genome survey sequence.
 SOURCE Brassica rapa subsp. pekinensis
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
 1 Bancroft, I., Huckle, E.J., Trick, M. and Beckett, P.
 TITLE Unpublished
 JOURNAL
 AUTHORS Beckett, P.
 TITLE Direct Submission
 JOURNAL Submitted (05-NOV-2004) Beckett P., Computational Biology, John
 Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk, NR4
 7UH, UNITED KINGDOM
 CONTACT: Ian Bancroft
 John Innes Centre Norwich Research Park, Colney, Norwich, NR4
 7UH, Tel: +44 1603 450843
 Fax: +44 1603 450821
 Email: ian.bancroft@bbsrc.ac.uk
 BAC End sequence of Brassica rapa BAC clone KBRH053P05 Seq primer:
 M13(forward)
 Class: BAC ends.

FEATURES
 source
 1..1338
 /organism="Brassica rapa subsp. pekinensis"
 /mol_type="genomic DNA"
 /cultivar="Chifu"
 /sub_species="pekinensis"
 /db_xref="taxon:51351"

/clone="KBRH053P05"
 /clone_11b="Brassica rapa BAC library (KBRH), Vector:
 pCUGBac1; Site 1: HindIII"

ORIGIN

Query Match 81.6%; Score 20.4; DB 10; Length 1338;
 Best Local Similarity 95.5%; Pred. No. 4.9e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGAT 22
 492 AAAAAAAAAATAAGCTTGAT 471

RESULT 49
 BG603017 160 bp mRNA linear EST 14-AUG-2001
 LOCUS BG603017
 DEFINITION EST502107 Plasmodium yoelii sporozoite cDNA Plasmodium yoelii cDNA
 clone PYCDN07, mRNA sequence.
 ACCESSION BG603017.1 GI:15153031
 VERSION
 KEYWORDS EST.
 SOURCE Plasmodium yoelii
 ORGANISM Plasmodium yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE
 1 (bases 1 to 160)
 Kappe, S.H.I., Gardner, M.J., Brown, S.M., Ross, J., Matuschewski, K.,
 Ribeiro, J.M., Adams, J.H., Quackembush, J., Cho, J., Carucci, D.J.,
 Hoffman, S.L. and Nussenzweig, V.
 Exploring the transcriptome of the malaria sporozoite stage
 Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9895-9900 (2001)

JOURNAL
 PUBLISHED 11493695

COMMENT Contact: Malcolm J. Gardner
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301 838 3519
 Fax: 301 838 0208
 Email: gardner@tigr.org
 Request for clones, please contact: Stefan Kappe,
 kappe01@popmail.med.nyu.edu Michael Heidelberg Division,
 Department of Pathology New York University School of Medicine.

FEATURES
 source
 1..160
 /organism="Plasmodium yoelii"
 /mol_type="mRNA"
 /strain="17XNL"
 /db_xref="taxon:5861"
 /clone="PYCDN07"
 /dev_stage="sporozoites from salivary gland"
 /lab_host="E. coli TOP10"
 /clone_11b="Plasmodium yoelii sporozoite cDNA"
 /note="Vector: PCR4; TA cloning; Plasmodium yoelii
 sporozoite cDNA library from salivary gland sporozoites 14
 days post-infection"

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 160;
 Best Local Similarity 88.0%; Pred. No. 5.9e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 25
 68 AAAAAAAAAATAAGTTGATCTT 92

RESULT 50
 AZ510748 206 bp DNA linear GSS 05-OCT-2000
 LOCUS AZ510748
 DEFINITION IM0355124F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0355124 F, genomic survey sequence.
 ACCESSION AZ510748
 VERSION AZ510748.1 GI:10692064

KEYWORDS

SOURCE

ORGANISM

GSS.
Mus musculus (house mouse)

REFERENCE

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 206)

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, B., Pedersen, T.,
Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von
Niederhausern, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

TITLE

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

Insert Length: 10000. Std Error: 0.00
Plate: 0355 row: 1 column: 24
Seq primer: CGTTGTAAGACGAGCCAGT
Class: plasmid ends
High quality sequence stop: 206.

FEATURES

Source

Location/Qualifiers
1..206
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCIM0355124"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_1lb="Mouse 10kb plasmid UGCIM library"
/note="Vector: FMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 206;
Best Local Similarity 88.0%; Pred. No. 5.9e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTT 25
|||||
Db 136 AAAAAAAAAAGCCTGATCTT 160
|||||

Search completed: December 14, 2005, 07:34:47
Job time: 1762.1 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:43:33 ; Search time 180.2 Seconds
(without alignments)
68.002 Million cell updates/sec

Title: US-10-681-773-3

Perfect score: 25
Sequence: 1 aaaaaaaaaaactgaagctgattc 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4161359 seqs, 24507644 residues

Total number of hits satisfying chosen parameters: 8322718

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications NA New:
1: /cgn2_6/ptoddata/1/pubpna/us09_NEW_PUB_seq.*
2: /cgn2_6/ptoddata/1/pubpna/us06_NEW_PUB_seq.*
3: /cgn2_6/ptoddata/1/pubpna/us07_NEW_PUB_seq.*
4: /cgn2_6/ptoddata/1/pubpna/us08_NEW_PUB_seq.*
5: /cgn2_6/ptoddata/1/pubpna/PCR_NEW_PUB_seq.*
6: /cgn2_6/ptoddata/1/pubpna/US10_NEW_PUB_seq.*
7: /cgn2_6/ptoddata/1/pubpna/US11_NEW_PUB_seq.*
8: /cgn2_6/ptoddata/1/pubpna/US11_NEW_PUB_seq.*
9: /cgn2_6/ptoddata/1/pubpna/US11_NEW_PUB_seq.*
10: /cgn2_6/ptoddata/1/pubpna/us60_NEW_PUB_seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	80.8	130733	7 US-11-121-086-19	Sequence 19, Appl
2	20.2	80.8	171732	7 US-11-121-086-98	Sequence 98, Appl
3	20.2	80.8	189539	7 US-11-121-086-16	Sequence 16, Appl
4	19.4	77.6	1433	6 US-10-750-185-40315	Sequence 40315, A
5	19.2	76.8	201	6 US-10-995-561-28493	Sequence 28493, A
6	19.2	76.8	905	6 US-10-750-185-54504	Sequence 54504, A
7	19.2	76.8	218821	7 US-11-121-086-31	Sequence 31, Appl
8	19.2	76.8	222094	6 US-10-995-561-13244	Sequence 13244, A
9	18.6	74.4	201	6 US-10-995-561-62773	Sequence 62773, A
10	18.6	74.4	1069	6 US-10-750-185-82543	Sequence 82543, A
11	18.6	74.4	1331	6 US-10-750-185-52544	Sequence 52544, A
12	18.6	74.4	1331	6 US-10-750-185-34884	Sequence 34884, A
13	18.6	74.4	1334	6 US-10-750-185-62604	Sequence 62604, A
14	18.6	74.4	1597	6 US-10-750-185-42069	Sequence 42069, A
15	18.6	74.4	1658	6 US-10-750-185-33931	Sequence 33931, A
16	18.6	74.4	1868	6 US-10-750-185-30064	Sequence 30064, A
17	18.6	74.4	1872	6 US-10-750-185-46372	Sequence 46372, A
18	18.6	74.4	2125	6 US-10-750-185-41385	Sequence 41385, A
19	18.6	74.4	2237	6 US-10-750-185-40001	Sequence 40001, A
20	18.6	74.4	2251	6 US-10-750-185-36679	Sequence 36679, A
21	18.6	74.4	2843	6 US-10-750-185-27339	Sequence 27339, A
22	18.6	74.4	4399	6 US-10-750-185-52104	Sequence 52104, A
23	18.6	74.4	27240	6 US-10-995-561-13499	Sequence 13499, A

24	18.6	74.4	52530	6 US-10-995-561-13360	Sequence 13360, A
25	18.6	74.4	159146	7 US-11-121-086-49	Sequence 49, Appl
26	18.6	74.4	132395	6 US-10-995-561-13277	Sequence 13277, A
27	18.6	74.4	182303	7 US-11-121-086-45	Sequence 45, Appl
28	18.6	74.4	191350	6 US-10-857-780-4	Sequence 4, Appl
29	18.6	74.4	260209	6 US-10-933-025-23	Sequence 23, Appl
30	18.6	74.4	340000	7 US-11-102-978-3	Sequence 3, Appl
31	18.6	74.4	387780	6 US-10-995-561-13259	Sequence 13259, A
32	18.4	73.6	201	6 US-10-995-561-47669	Sequence 47669, A
33	18.4	73.6	645179	6 US-10-750-185-33726	Sequence 13293, A
34	18.2	72.8	1009	6 US-10-750-185-33776	Sequence 33776, A
35	18.2	72.8	1605	6 US-10-750-185-59676	Sequence 59676, A
36	18.2	72.8	1766	6 US-10-750-185-36006	Sequence 36006, A
37	18.2	72.8	1925	6 US-10-750-185-57026	Sequence 57026, A
38	18.2	72.8	50353	6 US-10-995-561-13305	Sequence 13305, A
39	18.2	72.8	188682	7 US-11-112-908-23	Sequence 23, Appl
40	18	72.0	1728	6 US-10-750-185-50634	Sequence 50634, A
41	17.8	71.2	420	6 US-10-927-641-62	Sequence 62, Appl
42	17.8	71.2	1599	6 US-10-750-185-47179	Sequence 47179, A
43	17.8	71.2	1648	6 US-10-750-185-42001	Sequence 42001, A
44	17.8	71.2	2688	6 US-10-750-185-55769	Sequence 55769, A
45	17.8	71.2	3720	6 US-10-927-641-127	Sequence 127, App
46	17.8	71.2	6450	7 US-11-091-668-3	Sequence 3, Appl
47	17.8	71.2	96539	6 US-10-995-561-13289	Sequence 13289, A
48	17.8	71.2	101001	6 US-10-995-561-13285	Sequence 13285, A
49	17.8	71.2	126552	7 US-11-121-086-1	Sequence 1, Appl
50	17.8	71.2	172543	7 US-11-121-086-6	Sequence 6, Appl
51	17.8	71.2	191684	7 US-11-121-086-2	Sequence 2, Appl
52	17.8	71.2	193084	7 US-11-121-086-82	Sequence 82, Appl
53	17.6	70.4	201	6 US-10-995-561-31568	Sequence 31568, A
54	17.6	70.4	201	6 US-10-995-561-31640	Sequence 31640, A
55	17.6	70.4	201	6 US-10-995-561-31641	Sequence 31641, A
56	17.6	70.4	201	6 US-10-995-561-82062	Sequence 82062, A
57	17.6	70.4	600	6 US-10-750-185-2657	Sequence 2657, Ap
58	17.6	70.4	600	6 US-10-750-185-21354	Sequence 21354, A
59	17.6	70.4	1145	6 US-10-750-185-46768	Sequence 46768, A
60	17.6	70.4	1431	6 US-10-750-185-26879	Sequence 26879, A
61	17.6	70.4	1431	6 US-10-750-185-49713	Sequence 49713, A
62	17.6	70.4	1520	6 US-10-750-185-55886	Sequence 55886, A
63	17.6	70.4	1680	6 US-10-750-185-57535	Sequence 57535, A
64	17.6	70.4	1754	6 US-10-750-185-58714	Sequence 58714, A
65	17.6	70.4	2058	6 US-10-750-185-47588	Sequence 47588, A
66	17.6	70.4	2030	6 US-10-750-185-40402	Sequence 40402, A
67	17.6	70.4	2474	6 US-10-750-185-25261	Sequence 25261, A
68	17.6	70.4	2490	6 US-10-750-185-54317	Sequence 54317, A
69	17.6	70.4	2713	6 US-10-750-185-32766	Sequence 32766, A
70	17.6	70.4	2725	6 US-10-750-185-59886	Sequence 59886, A
71	17.6	70.4	4335	6 US-10-523-912-3	Sequence 3, Appl
72	17.6	70.4	4337	6 US-10-523-912-1	Sequence 1, Appl
73	17.6	70.4	40349	7 US-11-117-187-184	Sequence 184, Appl
74	17.6	70.4	55826	6 US-10-995-561-13356	Sequence 13356, A
75	17.6	70.4	58682	7 US-11-117-187-205	Sequence 205, App
76	17.6	70.4	98776	6 US-10-995-561-13331	Sequence 13331, A
77	17.6	70.4	119036	6 US-10-995-561-13314	Sequence 13314, A
78	17.6	70.4	149111	7 US-11-112-908-63	Sequence 63, Appl
79	17.6	70.4	156297	7 US-11-121-086-65	Sequence 65, Appl
80	17.6	70.4	157230	7 US-11-112-908-64	Sequence 64, Appl
81	17.6	70.4	160226	7 US-11-112-908-62	Sequence 62, Appl
82	17.6	70.4	170508	7 US-11-112-908-62	Sequence 62, Appl
83	17.6	70.4	189252	7 US-11-121-086-54	Sequence 54, Appl
84	17.6	70.4	190882	7 US-11-121-086-69	Sequence 69, Appl
85	17.6	70.4	191797	7 US-11-121-086-13	Sequence 13, Appl
86	17.6	70.4	611587	7 US-11-117-187-209	Sequence 209, App
87	17.6	70.4	1082144	7 US-11-117-187-211	Sequence 211, App
88	17.4	69.6	1599	6 US-10-750-185-28517	Sequence 28517, A
89	17.4	69.6	1645	6 US-10-750-185-42555	Sequence 42555, A
90	17.4	69.6	212716	7 US-11-121-086-95	Sequence 95, Appl
91	17.4	69.6	340000	7 US-11-102-978-3	Sequence 3, Appl
92	17.2	68.8	201	6 US-10-995-561-25339	Sequence 25339, A
93	17.2	68.8	201	6 US-10-995-561-25241	Sequence 25241, A
94	17.2	68.8	201	6 US-10-995-561-58454	Sequence 58454, A
95	17.2	68.8	201	6 US-10-995-561-69364	Sequence 69364, A
96	17.2	68.8	201	6 US-10-995-561-78764	Sequence 78764, A

```

c 97 17.2 68.8 201 6 US-10-995-561-80597 Sequence 80597, A
c 98 17.2 68.8 201 6 US-10-995-561-81867 Sequence 81867, A
c 99 17.2 68.8 201 6 US-10-995-561-82231 Sequence 82231, A
c 100 17.2 68.8 201 6 US-10-995-561-82599 Sequence 82599, A
c 101 17.2 68.8 201 6 US-10-995-561-83005 Sequence 83005, A
c 102 17.2 68.8 201 6 US-10-995-561-83307 Sequence 83307, A
c 103 17.2 68.8 201 6 US-10-995-561-83618 Sequence 83618, A
c 104 17.2 68.8 201 6 US-10-995-561-83932 Sequence 83932, A
c 105 17.2 68.8 201 6 US-10-995-561-84244 Sequence 84244, A
c 106 17.2 68.8 201 6 US-10-995-561-84567 Sequence 84567, A
c 107 17.2 68.8 201 6 US-10-995-561-84890 Sequence 84890, A
c 108 17.2 68.8 201 6 US-10-995-561-85213 Sequence 85213, A
c 109 17.2 68.8 201 6 US-10-995-561-85521 Sequence 85521, A
c 110 17.2 68.8 201 6 US-10-995-561-85833 Sequence 85833, A
c 111 17.2 68.8 201 6 US-10-995-561-86145 Sequence 86145, A
c 112 17.2 68.8 201 6 US-10-995-561-86457 Sequence 86457, A
c 113 17.2 68.8 201 6 US-10-995-561-86769 Sequence 86769, A
c 114 17.2 68.8 201 6 US-10-995-561-87081 Sequence 87081, A
c 115 17.2 68.8 201 6 US-10-995-561-87393 Sequence 87393, A
c 116 17.2 68.8 201 6 US-10-995-561-87705 Sequence 87705, A
c 117 17.2 68.8 201 6 US-10-995-561-88017 Sequence 88017, A
c 118 17.2 68.8 201 6 US-10-995-561-88329 Sequence 88329, A
c 119 17.2 68.8 201 6 US-10-995-561-88641 Sequence 88641, A
c 120 17.2 68.8 201 6 US-10-995-561-88953 Sequence 88953, A
c 121 17.2 68.8 201 6 US-10-995-561-89265 Sequence 89265, A
c 122 17.2 68.8 201 6 US-10-995-561-89577 Sequence 89577, A
c 123 17.2 68.8 201 6 US-10-995-561-89889 Sequence 89889, A
c 124 17.2 68.8 201 6 US-10-995-561-90201 Sequence 90201, A
c 125 17.2 68.8 201 6 US-10-995-561-90513 Sequence 90513, A
c 126 17.2 68.8 201 6 US-10-995-561-90825 Sequence 90825, A
c 127 17.2 68.8 201 6 US-10-995-561-91137 Sequence 91137, A
c 128 17.2 68.8 201 6 US-10-995-561-91449 Sequence 91449, A
c 129 17.2 68.8 201 6 US-10-995-561-91761 Sequence 91761, A
c 130 17.2 68.8 201 6 US-10-995-561-92073 Sequence 92073, A
c 131 17.2 68.8 201 6 US-10-995-561-92385 Sequence 92385, A
c 132 17.2 68.8 201 6 US-10-995-561-92697 Sequence 92697, A
c 133 17.2 68.8 201 6 US-10-995-561-93009 Sequence 93009, A
c 134 17.2 68.8 201 6 US-10-995-561-93321 Sequence 93321, A
c 135 17.2 68.8 201 6 US-10-995-561-93633 Sequence 93633, A
c 136 17.2 68.8 201 6 US-10-995-561-93945 Sequence 93945, A
c 137 17.2 68.8 201 6 US-10-995-561-94257 Sequence 94257, A
c 138 17.2 68.8 201 6 US-10-995-561-94569 Sequence 94569, A
c 139 17.2 68.8 201 6 US-10-995-561-94881 Sequence 94881, A
c 140 17.2 68.8 201 6 US-10-995-561-95193 Sequence 95193, A
c 141 17.2 68.8 201 6 US-10-995-561-95505 Sequence 95505, A
c 142 17.2 68.8 201 6 US-10-995-561-95817 Sequence 95817, A
c 143 17.2 68.8 201 6 US-10-995-561-96129 Sequence 96129, A
c 144 17.2 68.8 201 6 US-10-995-561-96441 Sequence 96441, A
c 145 17.2 68.8 201 6 US-10-995-561-96753 Sequence 96753, A
c 146 17.2 68.8 201 6 US-10-995-561-97065 Sequence 97065, A
c 147 17.2 68.8 201 6 US-10-995-561-97377 Sequence 97377, A
c 148 17.2 68.8 201 6 US-10-995-561-97689 Sequence 97689, A
c 149 17.2 68.8 201 6 US-10-995-561-98001 Sequence 98001, A
c 150 17.2 68.8 201 6 US-10-995-561-98313 Sequence 98313, A

```

ALIGNMENTS

```

RESULT 1
US-11-121-086-19
; Sequence 19, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107

```

```

; SOFTWARE: Patentin version 3.3
; SEQ ID NO 19
; LENGTH: 130733
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-19
Query Match 80.8%; Score 20.2; DB 7; Length 130733;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAACTTAAGCTTGATCTT 25
Db 49398 AAAAAAAAAATGAATCTTGATCTT 49422

```

```

RESULT 2
US-11-121-086-98
; Sequence 98, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 98
; LENGTH: 171732
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-98
Query Match 80.8%; Score 20.2; DB 7; Length 171732;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAACTTAAGCTTGATCTT 25
Db 120361 AAAAAAAAAATTAAGCTTGATCTT 120385

```

```

RESULT 3
US-11-121-086-16
; Sequence 16, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 16
; LENGTH: 189539
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-16
Query Match 80.8%; Score 20.2; DB 7; Length 189539;
Best Local Similarity 88.0%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAACTTAAGCTTGATCTT 25

```

Db 188341 AAAAAAAAAAATGAATCTTGATCTT 188365

RESULT 4

US-10-750-185-40315
 ; Sequence 40315, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 40315
 ; LENGTH: 1433
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866881121145
 US-10-750-185-40315

Query Match 77.6%; Score 19.4; DB 6; Length 1433;
 Best Local Similarity 95.2%; Pred. No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAATGAATCTTGATCTT 21
 Db 183 AAAAAAAAAAATGAATCTTGATCTT 203

RESULT 5

US-10-995-561-28493
 ; Sequence 28493, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; FILE REFERENCE: CU001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 28493
 ; LENGTH: 201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-995-561-28493

Query Match 76.8%; Score 19.2; DB 6; Length 201;
 Best Local Similarity 87.5%; Pred. No. 94;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAATGAATCTTGATCTT 24
 Db 94 AAAAAAAAAAGATGAAGCTTGATCTT 117

RESULT 6

US-10-750-185-54504/C
 ; Sequence 54504, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.

APPLICANT: DENISE, Sue K.
 APPLICANT: KERR, Richard
 APPLICANT: ROSENFELD, David
 APPLICANT: HOLM, Tom
 APPLICANT: BATES, Stephen
 APPLICANT: FANTIN, Dennis
 TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 FILE REFERENCE: MM1100-2
 CURRENT APPLICATION NUMBER: US/10/750,185
 CURRENT FILING DATE: 2003-12-31
 PRIOR APPLICATION NUMBER: US 60/437,482
 PRIOR FILING DATE: 2002-12-31
 NUMBER OF SEQ ID NOS: 64922
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 54504
 LENGTH: 905
 TYPE: DNA
 ORGANISM: Bovine 19866881526848
 US-10-750-185-54504

Query Match 76.8%; Score 19.2; DB 6; Length 905;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAATGAATCTTGATCTT 24
 Db 586 AAAAAAAAAAATGAATCTTGATCTT 563

RESULT 7

US-11-121-086-31
 ; Sequence 31, Application US/1121086
 ; Publication No. US2005026459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: POULSEN, TIM S.
 ; APPLICANT: NIELSEN, KIRSTEN V.
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
 ; FILE REFERENCE: 09138.6000-00000
 ; CURRENT APPLICATION NUMBER: US/11/121,086
 ; CURRENT FILING DATE: 2005-05-04
 ; PRIOR APPLICATION NUMBER: 60/567,570
 ; PRIOR FILING DATE: 2004-05-04
 ; NUMBER OF SEQ ID NOS: 107
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 31
 ; LENGTH: 218821
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens

FEATURE:
 NAME/KEY: modified_base
 LOCATION: (106949)..(106949)
 OTHER INFORMATION: a, c, g, t, unknown or other
 FEATURE:
 NAME/KEY: modified_base
 LOCATION: (110322)..(110324)
 OTHER INFORMATION: a, c, g, t, unknown or other
 FEATURE:
 NAME/KEY: modified_base
 LOCATION: (115133)..(115133)
 OTHER INFORMATION: a, c, g, t, unknown or other
 FEATURE:
 NAME/KEY: modified_base
 LOCATION: (131300)..(131300)
 OTHER INFORMATION: a, c, g, t, unknown or other
 FEATURE:
 NAME/KEY: modified_base
 LOCATION: (139059)..(139158)
 OTHER INFORMATION: a, c, g, t, unknown or other
 FEATURE:
 NAME/KEY: modified_base
 LOCATION: (157740)..(157740)
 OTHER INFORMATION: a, c, g, t, unknown or other
 FEATURE:

NAME/KEY: modified_base
LOCATION: (157777)..(157777)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157900)..(157900)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157919)..(157919)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157926)..(157926)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158094)..(158094)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158138)..(158138)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158193)..(158193)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158241)..(158242)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158259)..(158259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158278)..(158278)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158295)..(158295)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158740)..(158839)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158929)..(158929)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163528)..(163530)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163550)..(163550)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163765)..(163765)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164000)..(164000)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164047)..(164047)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base

LOCATION: (164084)..(164084)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167233)..(167233)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167236)..(167236)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167238)..(167238)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170247)..(170247)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170249)..(170250)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170252)..(170253)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170259)..(170259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170262)..(170263)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170266)..(170266)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174470)..(174470)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174472)..(174472)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174474)..(174474)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179059)..(179060)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179064)..(179064)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197001)..(197001)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197004)..(197005)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197007)..(197007)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (200349)..(200349)


```
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (200351)..(200351)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (212425)..(212426)
OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-31

Query Match      76.8%; Score 19.2; DB 7; Length 218821;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 157186 AAAAAAAAACTAAAGCTTGATTT 157209

RESULT 8
US-10-995-561-13244
; Sequence 13244, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13244
; LENGTH: 222094
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13244

Query Match      76.8%; Score 19.2; DB 6; Length 222094;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 24
Db 90938 AAAAAAAAAAGTAAGTTGATCT 90961

RESULT 9
US-10-995-561-62773
; Sequence 62773, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62773
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-62773

Query Match      74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAAGCTTGATCT 25
Db 1 AAAAAAAAAAAAAAAAAAGTTGACCTT 25

RESULT 10
US-10-995-561-82633/C
; Sequence 82633, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82633
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-82633

Query Match      74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCT 25
Db 201 AAAAAAAAAAAAAAAAAAGTTGACCTT 177

RESULT 11
US-10-750-185-52544/C
; Sequence 52544, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFIELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52544
; LENGTH: 1069
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-52544

Query Match      74.4%; Score 18.6; DB 6; Length 1069;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 53 AAAAAAAAAACCAAACTTGAAATTT 29

RESULT 12
US-10-750-185-34894/C
; Sequence 34894, Application US/10750185
; Publication No. US20050260603A1
```

```

; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34894
; LENGTH: 1331
; TYPE: DNA
; ORGANISM: Bovine 19866881153190
US-10-750-185-34894

Query Match          74.4%; Score 18.6; DB 6; Length 1331;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTAAGCTTGATCTT 25
Db 895 AAAAAAAAAAAGCTTAAGCTTGATCTT 871

RESULT 13
US-10-750-185-62604/c
; Sequence 62604, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62604
; LENGTH: 1354
; TYPE: DNA
; ORGANISM: Bovine 19866880368620
US-10-750-185-62604

Query Match          74.4%; Score 18.6; DB 6; Length 1354;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTAAGCTTGATCTT 25
Db 200 AAAAAAAAAAAGCTTAAGCTTGATCTT 176

RESULT 14
US-10-750-185-42089/c
; Sequence 42089, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David

```

```

; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42089
; LENGTH: 1597
; TYPE: DNA
; ORGANISM: Bovine 19866880543152
US-10-750-185-42089

Query Match          74.4%; Score 18.6; DB 6; Length 1597;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTAAGCTTGATCTT 25
Db 1375 AAAAAAAAAAAGCTTAAGCTTGATCTT 1351

RESULT 15
US-10-750-185-33931/c
; Sequence 33931, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33931
; LENGTH: 1658
; TYPE: DNA
; ORGANISM: Bovine 19866880421345
US-10-750-185-33931

Query Match          74.4%; Score 18.6; DB 6; Length 1658;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCTTAAGCTTGATCTT 25
Db 343 AAAAAAAAAAAGCTTAAGCTTGATCTT 319

RESULT 16
US-10-750-185-30064
; Sequence 30064, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom

```

```
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30064
; LENGTH: 1868
; TYPE: DNA
; ORGANISM: Bovine 19866881232408
US-10-750-185-30064

Query Match      74.4%; Score 18.6; DB 6; Length 1868;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 730 AAAAAAAAAAAAGCTTAATTT 754

RESULT 17
US-10-750-185-46372/C
; Sequence 46372, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46372
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Bovine 19866881061171
US-10-750-185-46372

Query Match      74.4%; Score 18.6; DB 6; Length 1872;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 187 AAAAAAAAAAACTAAAGATTTT 163

RESULT 18
US-10-750-185-41395
; Sequence 41395, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
```

```
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41395
; LENGTH: 2125
; TYPE: DNA
; ORGANISM: Bovine 19866880374636
US-10-750-185-41395

Query Match      74.4%; Score 18.6; DB 6; Length 2125;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 1886 AAAAAAAAAAAACAACTTCATCTT 1910

RESULT 19
US-10-750-185-40001
; Sequence 40001, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40001
; LENGTH: 2237
; TYPE: DNA
; ORGANISM: Bovine 19866880291279
US-10-750-185-40001

Query Match      74.4%; Score 18.6; DB 6; Length 2237;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 626 AAAAAAAAAAAAGCTTATCTT 650

RESULT 20
US-10-750-185-36679/C
; Sequence 36679, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
```

```

; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36679
; LENGTH: 2251
; TYPE: DNA
; ORGANISM: Bovine 19866880665289
US-10-750-185-36679

Query Match          74.4%; Score 18.6; DB 6; Length 2251;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 448 AACAAAAAAATTAATCTTGCTT 424

RESULT 21
US-10-750-185-27339
; Sequence 27339, Application US/10750185
; Publication No.: US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFIELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27339
; LENGTH: 2843
; TYPE: DNA
; ORGANISM: Bovine 19866880798157
US-10-750-185-27339

Query Match          74.4%; Score 18.6; DB 6; Length 2843;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 2673 AAAAAGAGAGTCAAGCTTGATCTT 2697

RESULT 22
US-10-750-185-52104
; Sequence 52104, Application US/10750185
; Publication No.: US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFIELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52104
; LENGTH: 4399
; TYPE: DNA
; ORGANISM: Bovine 19866880511075
US-10-750-185-52104

Query Match          74.4%; Score 18.6; DB 6; Length 4399;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 718 AAAAAAAAAAAAGATTGCTT 742

RESULT 23
US-10-995-561-13499/C
; Sequence 13499, Application US/10995561
; Publication No.: US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13499
; LENGTH: 27240
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13499

Query Match          74.4%; Score 18.6; DB 6; Length 27240;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 25
Db 24221 AAAAAAAAAAAAGTTGACCTT 24197

RESULT 24
US-10-995-561-13360
; Sequence 13360, Application US/10995561
; Publication No.: US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13360
; LENGTH: 52520
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(52520)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13360

Query Match          74.4%; Score 18.6; DB 6; Length 52520;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 43682 AAAAAAAAAAAAGTTTGACCTT 43706

RESULT 25
US-11-121-086-49/c
; Sequence 49, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 49
; LENGTH: 159146
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-49

Query Match 74.4%; Score 18.6; DB 7; Length 159146;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 96286 AAAAAAAAACTAGCATGATTTT 96262

RESULT 26
US-10-995-561-13277/c
; Sequence 13277, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CU001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13277
; LENGTH: 173995
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13277

Query Match 74.4%; Score 18.6; DB 6; Length 173995;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 71234 AAAAAACAACTCATGCTTATTTT 71210

RESULT 27
US-11-121-086-45
; Sequence 45, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
```

```
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 45
; LENGTH: 182303
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-45

Query Match 74.4%; Score 18.6; DB 7; Length 182303;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 25
Db 127101 AAAAAAAAAAAAGTTTATCTT 127125

RESULT 28
US-10-857-780-4
; Sequence 4, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENEALAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 191350
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (16914)..(16914)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (19266)..(19266)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (26334)..(26334)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34472)..(34472)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (38627)..(38628)
; OTHER INFORMATION: n is a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (40555)..(40555)
; OTHER INFORMATION: n is a, c, g, or t
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (57355)..(57355)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (61207)..(61207)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (64980)..(64980)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (64987)..(64987)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (81452)..(81452)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (95893)..(95893)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (103359)..(103359)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (117565)..(117565)
OTHER INFORMATION: n is a, c, g, or t
US-10-857-780-4
```

```
Query Match          74.4%; Score 18.6; DB 6; Length 191350;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCTT 25
```

```
Db 79676 AAAAAACAAACAAACCTTGATATT 79700
```

```
RESULT 29
US-10-933-025-23
Sequence 23, Application US/10933025
Publication No. US20050265987A1
GENERAL INFORMATION:
APPLICANT: ROSEN, STEVEN
APPLICANT: HEMMERICH, STEFAN
APPLICANT: TOMITA, MEGUMI
TITLE OF INVENTION: Sulfotransferases and methods of use
FILE REFERENCE: UCAL-230CON
CURRENT APPLICATION NUMBER: US/10/933,025
PRIOR FILING DATE: 2004-09-01
PRIOR APPLICATION NUMBER: 10/025,966
PRIOR FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/258,577
PRIOR FILING DATE: 2000-12-27
PRIOR APPLICATION NUMBER: 60/267,831
PRIOR FILING DATE: 2001-09-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 260209
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(260209)
OTHER INFORMATION: n = A,T,C or G
US-10-933-025-23
```

```
Query Match          74.4%; Score 18.6; DB 6; Length 260209;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCTT 25
```

```
Db 20986 AAAAAACAAAGCTTGATCTT 21010
```

```
RESULT 30
US-11-102-978-3
Sequence 3, Application US/11102978
Publication No. US20050250142A1
GENERAL INFORMATION:
APPLICANT: University of Utah Technology Transfer Office
APPLICANT: University of Utah Research Foundation
TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
FILE REFERENCE: 0274-5537, IUS
CURRENT APPLICATION NUMBER: US/11/102,978
PRIOR FILING DATE: 2005-04-11
PRIOR APPLICATION NUMBER: PCT/US2003/033152
PRIOR FILING DATE: 2003-10-18
PRIOR APPLICATION NUMBER: 60/419,576
PRIOR FILING DATE: 2002-10-18
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 340000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: (56948)..(57115)
OTHER INFORMATION: C21orf34 exon
FEATURE:
NAME/KEY: misc feature
LOCATION: (80056)..(81089)
OTHER INFORMATION: Gene VDAC2P, voltage-dependent anion channel isoform 2 pseudogen
FEATURE:
NAME/KEY: exon
LOCATION: (167308)..(167438)
OTHER INFORMATION: C21orf34 exon
FEATURE:
NAME/KEY: exon
LOCATION: (216732)..(216833)
OTHER INFORMATION: C21orf34 exon
US-11-102-978-3
```

```
Query Match          74.4%; Score 18.6; DB 7; Length 340000;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCTT 25
```

```
Db 151514 AAAAAACAAAGCTTGATCTT 151538
```

```
RESULT 31
US-10-995-561-13259
Sequence 13259, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CLO01559
CURRENT APPLICATION NUMBER: US/10/995,561
PRIOR FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13259
```

```
LENGTH: 387780
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(387780)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13259
```

```
Query Match 74.4%; Score 18.6; DB 6; Length 387780;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATCTT 25
DB 281125 AAAAAAAAAAAGTTTATCTT 281149
```

```
RESULT 32
US-10-995-561-47669
Sequence 47669, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: CL001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 47669
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-47669
```

```
Query Match 73.6%; Score 18.4; DB 6; Length 201;
Best Local Similarity 95.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 AAAAAAAAACTAAGCTTGAT 22
DB 58 AAAAAAAAAATTAAGCTTGAT 77
```

```
RESULT 33
US-10-995-561-13293
Sequence 13293, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: CL001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 13293
LENGTH: 645179
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-13293
```

```
Query Match 73.6%; Score 18.4; DB 6; Length 645179;
Best Local Similarity 95.0%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 AAAAAAAAACTAAGCTTGAT 22
DB 58 AAAAAAAAAATTAAGCTTGAT 77
```

```
DB 75327 AAAAAAAAAATTAAGCTTGAT 75346
```

```
RESULT 34
US-10-750-185-33776
Sequence 33776, Application US/10750185
Publication No. US2005026603A1
GENERAL INFORMATION:
APPLICANT: MMT GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MW1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIORITY APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33776
LENGTH: 1009
TYPE: DNA
ORGANISM: Bovine
US-10-750-185-33776
```

```
Query Match 72.8%; Score 18.2; DB 6; Length 1009;
Best Local Similarity 87.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATC 23
DB 626 AAAAAAAAAATTAAGCTTGTC 648
```

```
RESULT 35
US-10-750-185-59676/C
Sequence 59676, Application US/10750185
Publication No. US2005026603A1
GENERAL INFORMATION:
APPLICANT: MMT GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MW1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIORITY APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 59676
LENGTH: 1605
TYPE: DNA
ORGANISM: Bovine
US-10-750-185-59676
```

```
Query Match 72.8%; Score 18.2; DB 6; Length 1605;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAAGCTTGATC 23
DB 905 AAAAAAAAAATTAAGCATGATC 883
```

```
RESULT 36
US-10-750-185-36006
; Sequence 36006, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFEUD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36006
; LENGTH: 1756
; TYPE: DNA
; ORGANISM: Bovine 19866881236123
US-10-750-185-36006

Query Match          72.8%; Score 18.2; DB 6; Length 1756;
Best Local Similarity 87.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAAAACTAAGCTTGATCT 25
DB 972 AAAAAAAAACTAAGCTTAATTT 994

RESULT 37
US-10-750-185-57026/c
; Sequence 57026, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFEUD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57026
; LENGTH: 1925
; TYPE: DNA
; ORGANISM: Bovine 19866880959916
US-10-750-185-57026

Query Match          72.8%; Score 18.2; DB 6; Length 1925;
Best Local Similarity 87.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATC 23
DB 1175 AAAAAAGTAACAAAGCTTGATC 1153

RESULT 38
US-10-995-561-13305
; Sequence 13305, Application US/10995561
```

```
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; PRIOR FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13305
; LENGTH: 50353
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13305

Query Match          72.8%; Score 18.2; DB 6; Length 50353;
Best Local Similarity 87.0%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCT 24
DB 49495 AAAAAAAATTAAGGTTGATCT 49517

RESULT 39
US-11-112-908-23
; Sequence 23, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 23
; LENGTH: 188682
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-23

Query Match          72.8%; Score 18.2; DB 7; Length 188682;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATC 23
DB 124820 AAAAAAAATTAATCTTGATC 124842

RESULT 40
US-10-750-185-50634/c
; Sequence 50634, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFEUD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
```



```

; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50634
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Bovine 19866881292087
US-10-750-185-50634

Query Match
Best Local Similarity 100.0%; Score 18; DB 6; Length 1728;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCT 18
Db 196 AAAAAAAAACTAAGCT 179

RESULT 41
US-10-927-641-62
; Sequence 62, Application US/10927641
; Publication No. US20050244968A1
; GENERAL INFORMATION:
; APPLICANT: Pereira, Ranjan
; APPLICANT: Rice, Stephen
; APPLICANT: Bagleton, Clare
; APPLICANT: Lasham, Annette
; APPLICANT: Wood, Marion
; APPLICANT: Visser, Elizabeth
; TITLE OF INVENTION: Modifications and Methods for the
; FILE REFERENCE: 11000.10364
; CURRENT APPLICATION NUMBER: US/10/927,641
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/10/137,036
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: PCT/NZ 01/00115
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: U.S. No. 09/724,624
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 09/598,401
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: PCT/NZ00/00018
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: U.S. No. 60/146,591
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: U.S. No. 09/276,599
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-927-641-62

Query Match
Best Local Similarity 71.2%; Score 17.8; DB 6; Length 420;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGA 21
Db 229 AAAAAAAAACTAAGCTGA 249

RESULT 42
US-10-750-185-47179
```

```

; Sequence 47179, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47179
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: Bovine 19866880818333
US-10-750-185-47179

Query Match
Best Local Similarity 71.2%; Score 17.8; DB 6; Length 1599;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGA 21
Db 1006 AAAAAAAAACTAAGCTGA 1026

RESULT 43
US-10-750-185-42001/C
; Sequence 42001, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42001
; LENGTH: 1648
; TYPE: DNA
; ORGANISM: Bovine 19866880861972
US-10-750-185-42001

Query Match
Best Local Similarity 71.2%; Score 17.8; DB 6; Length 1648;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGA 21
Db 685 AAAAAAAAACTAAGCTGA 665

RESULT 44
US-10-750-185-55769/C
; Sequence 55769, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
```

```

; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55769
; LENGTH: 2688
; TYPE: DNA
; ORGANISM: Bovine 1986680540328
US-10-750-185-55769

Query Match          71.2%; Score 17.8; DB 6; Length 2688;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGA 21
DB 430 AAAAAAAAAATTAAGCTTGA 410

RESULT 45
US-10-927-641-127
; Sequence 127, Application US/10927641
; GENERAL INFORMATION:
; APPLICANT: Perera, Ranjan
; APPLICANT: Rice, Stephen
; APPLICANT: Bagleton, Clare
; APPLICANT: Lasham, Annette
; APPLICANT: Wood, Marion
; APPLICANT: Vasser, Elizabeth
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Expression
; FILE REFERENCE: 11000.1036c4
; CURRENT APPLICATION NUMBER: US/10/927,641
; PRIOR FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/10/137,036
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: PCT/NZ 01/00115
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: U.S. No. 09/724,624
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 09/598,401
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: PCT/NZ00/00018
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: U.S. No. 60/146,591
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: U.S. No. 09/276,599
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127
; LENGTH: 3720
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-927-641-127

Query Match          71.2%; Score 17.8; DB 6; Length 3720;
Best Local Similarity 90.5%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGA 21
```

```

DB 3313 AAAAAAAAAAAAAAAAAAGCTTGA 3333

RESULT 46
US-11-091-668-3/c
; Sequence 3, Application US/11091668
; Publication No. US20050262585A1
; GENERAL INFORMATION:
; APPLICANT: University of Nebraska
; APPLICANT: Mackenzie, Sally Ann
; APPLICANT: Vachchippavala, Zarir Erach
; TITLE OF INVENTION: Soybean PAM Synthase Promoters Useful in Parasite Control
; FILE REFERENCE: 1231-221
; CURRENT APPLICATION NUMBER: US/11/091,668
; CURRENT FILING DATE: 2005-03-28
; PRIOR APPLICATION NUMBER: 60556745
; PRIOR FILING DATE: 2004-03-26
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 6450
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2484)..(6425)
US-11-091-668-3

Query Match          71.2%; Score 17.8; DB 7; Length 6450;
Best Local Similarity 90.5%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGA 21
DB 2017 AAAAAAAAAAAAAAAAAAGCTTGA 1997

RESULT 47
US-10-995-561-13289
; Sequence 13289, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13289
; LENGTH: 96539
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13289

Query Match          71.2%; Score 17.8; DB 6; Length 96539;
Best Local Similarity 90.5%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGA 21
DB 94332 AAAAAAAAAACAAAGCTTGA 94352

RESULT 48
US-10-995-561-13255/c
; Sequence 13255, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
```

```
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13255
; LENGTH: 101001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(101001)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13255

Query Match          71.2%; Score 17.8; DB 6; Length 101001;
Best Local Similarity 90.5%; Pred. No. 6.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAAGCTGA 21
Db      86638 AAAAAAAAACTAAAGCTTA 86618

RESULT 49
US-11-121-086-1
; Sequence 1, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 126552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-1

Query Match          71.2%; Score 17.8; DB 7; Length 126552;
Best Local Similarity 90.5%; Pred. No. 7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAAGCTGA 21
Db      99187 AAAAAAAAACTAAAGCTGA 99207

RESULT 50
US-11-121-086-6/c
; Sequence 6, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6
```

```
; LENGTH: 172543
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-6

Query Match          71.2%; Score 17.8; DB 7; Length 172543;
Best Local Similarity 90.5%; Pred. No. 7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAAGCTGA 21
Db      126313 AAAAAAAAACTAAAGCTGA 126293

Search completed: December 14, 2005, 11:40:18
Job time : 188.2 secs
```

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-4

Perfect score: 25

Sequence: 1 aaaaaaaaaactagctgacatc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : GenEmbl :*

1: gb_ba :*

2: gb_in :*

3: gb_env :*

4: gb_om :*

5: gb_ov :*

6: gb_pat :*

7: gb_ph :*

8: gb_pr :*

9: gb_ro :*

10: gb_sts :*

11: gb_sy :*

12: gb_un :*

13: gb_vi :*

14: gb_hcg :*

15: gb_pl :*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	147404	8	AC004738 Homo sapi
2	22.4	89.6	28143	8	AC124305 Homo sapi
3	22.4	89.6	142745	8	AC124311 Homo sapi
4	22.4	89.6	179810	8	AC124997 Homo sapi
5	21.8	87.2	51096	14	AC165510 Homo sapi
6	21.8	87.2	214779	14	AC112769 Homo sapi
7	21.8	87.2	248019	14	AC109007 Homo sapi
8	21.4	85.6	140150	8	AC072057 Homo sapi
9	20.8	83.2	724	10	BV456584 Homo sapi
10	20.8	83.2	787	10	BV491883 Homo sapi
11	20.8	83.2	787	10	BV536707 Homo sapi
12	20.8	83.2	837	10	BV506331 Homo sapi
13	20.8	83.2	87782	14	AP008030 Homo sapi
14	20.8	83.2	88311	14	AL358954 Homo sapi
15	20.8	83.2	96008	14	AP007366 Homo sapi
16	20.8	83.2	110000	15	AP008208 Homo sapi
17	20.8	83.2	115335	8	AC093811 Homo sapi
18	20.8	83.2	132282	15	AP004855 Homo sapi

19	20.8	83.2	137278	9	AC104103 Homo sapi
20	20.8	83.2	169780	8	AL353784 Homo sapi
21	20.8	83.2	184554	8	BS000168 Homo sapi
22	20.8	83.2	196721	8	AC019179 Homo sapi
23	20.8	83.2	200721	8	AC105245 Homo sapi
24	20.8	83.2	220407	14	AC121626 Homo sapi
25	20.8	83.2	222358	14	AC099385 Homo sapi
26	20.8	83.2	222669	14	AC125295 Homo sapi
27	20.8	83.2	223132	9	AC123870 Homo sapi
28	20.8	83.2	228995	14	AC129388 Homo sapi
29	20.8	83.2	248526	14	AC095612 Homo sapi
30	20.8	83.2	249383	14	AC103260 Homo sapi
31	20.8	83.2	255246	14	AC108639 Homo sapi
32	20.8	83.2	258727	14	AC103272 Homo sapi
33	20.8	83.2	281349	14	CR847855 Homo sapi
34	20.4	81.6	68644	14	AC083784 Homo sapi
35	20.4	81.6	68644	14	AC083784 Homo sapi
36	20.4	81.6	110000	9	AE014175_2 Homo sapi
37	20.4	81.6	136403	8	AC096773 Homo sapi
38	20.4	81.6	161391	8	AC021878 Homo sapi
39	20.4	81.6	164968	14	AC022939 Homo sapi
40	20.4	81.6	169557	8	AC008529 Homo sapi
41	20.4	81.6	201412	14	CT025554 Homo sapi
42	20.2	80.8	5228	6	AX346556 Homo sapi
43	20.2	80.8	6584	14	AC014427 Homo sapi
44	20.2	80.8	15649	6	AX348975 Homo sapi
45	20.2	80.8	18525	2	AY821561 Homo sapi
46	20.2	80.8	18525	2	AY821561 Homo sapi
47	20.2	80.8	65059	14	AC067765 Homo sapi
48	20.2	80.8	65091	14	AC100199 Homo sapi
49	20.2	80.8	84322	15	AP004509 Homo sapi
50	20.2	80.8	99596	8	AC106873 Homo sapi
51	20.2	80.8	102568	14	AP007491 Homo sapi
52	20.2	80.8	102619	8	HS69P17 Homo sapi
53	20.2	80.8	102716	14	AC161353 Homo sapi
54	20.2	80.8	110000	14	AC099448_2 Homo sapi
55	20.2	80.8	110000	14	AC140815_3 Homo sapi
56	20.2	80.8	110000	14	AC141302_2 Homo sapi
57	20.2	80.8	110000	15	AP008215_040 Homo sapi
58	20.2	80.8	110000	15	AP008215_040 Homo sapi
59	20.2	80.8	110000	15	AP008215_040 Homo sapi
60	20.2	80.8	113066	6	AC026729 Homo sapi
61	20.2	80.8	113452	15	AC135801 Homo sapi
62	20.2	80.8	120090	14	AC162130 Homo sapi
63	20.2	80.8	122889	8	AC012032 Homo sapi
64	20.2	80.8	123814	8	AC005393 Homo sapi
65	20.2	80.8	124379	14	AC144420 Homo sapi
66	20.2	80.8	124531	8	HS633H17 Homo sapi
67	20.2	80.8	125482	14	AC166513 Homo sapi
68	20.2	80.8	126872	14	AC142506 Homo sapi
69	20.2	80.8	134089	8	AC112917 Homo sapi
70	20.2	80.8	134674	15	AC137820 Homo sapi
71	20.2	80.8	134773	14	AC012347 Homo sapi
72	20.2	80.8	134774	8	AC023520 Homo sapi
73	20.2	80.8	136116	15	AC148762 Homo sapi
74	20.2	80.8	136271	5	CR385075 Homo sapi
75	20.2	80.8	137439	5	CR550308 Homo sapi
76	20.2	80.8	143657	14	AC156080 Homo sapi
77	20.2	80.8	144475	14	AC016785 Homo sapi
78	20.2	80.8	145722	14	AC015503 Homo sapi
79	20.2	80.8	146708	8	AL139375 Homo sapi
80	20.2	80.8	148743	5	BX324154 Homo sapi
81	20.2	80.8	151640	9	AC122053 Homo sapi
82	20.2	80.8	154828	8	AC069070 Homo sapi
83	20.2	80.8	155764	8	AC011238 Homo sapi
84	20.2	80.8	160936	14	AC073494 Homo sapi
85	20.2	80.8	162223	8	AC145988 Homo sapi
86	20.2	80.8	163958	2	AC010705 Homo sapi
87	20.2	80.8	164142	14	AC010152 Homo sapi
88	20.2	80.8	164276	8	AL354937 Homo sapi
89	20.2	80.8	164913	8	AP005359 Homo sapi
90	20.2	80.8	165196	14	AC140861 Homo sapi
91	20.2	80.8	165257	5	AL645790 Homo sapi

C 92	20.2	80.8	166019	14	AC015648	AC015648 Homo sapi
C 93	20.2	80.8	166451	14	AC136430	AC136430 Homo sapi
C 94	20.2	80.8	166909	14	AC140708	AC140708 Homo sapi
C 95	20.2	80.8	167060	14	AC141296	AC141296 Homo sapi
C 96	20.2	80.8	167784	14	AC163242	AC163242 Callacebu
C 97	20.2	80.8	168874	15	AP005738	AP005738 Oryza sat
C 98	20.2	80.8	168931	14	AC144998	AC144998 Pan trogl
C 99	20.2	80.8	169074	14	AC148830	AC148830 Pan trogl
C 100	20.2	80.8	169908	14	AC150910	AC150910 Pan trogl
C 101	20.2	80.8	170001	8	HS45P21	AL021917 Human DNA
C 102	20.2	80.8	171502	14	AC142382	AC142382 Homo sapi
C 103	20.2	80.8	171863	8	AP000459	AP000459 Homo sapi
C 104	20.2	80.8	172361	14	AC140857	AC140857 Homo sapi
C 105	20.2	80.8	173692	14	AC027226	AC027226 Homo sapi
C 106	20.2	80.8	174215	14	AC144417	AC144417 Rattus no
C 107	20.2	80.8	174507	9	AC132955	AC132955 Mus muscu
C 108	20.2	80.8	174662	14	AC026036	AC026036 Homo sapi
C 109	20.2	80.8	175106	14	AC141606	AC141606 Homo sapi
C 110	20.2	80.8	175880	8	BX001068	BX001068 Mouse DNA
C 111	20.2	80.8	176872	8	AC012078	AC012078 Homo sapi
C 112	20.2	80.8	178410	14	AC145190	AC145190 Sus scrof
C 113	20.2	80.8	178461	14	AC141059	AC141059 Homo sapi
C 114	20.2	80.8	179083	14	AC141281	AC141281 Homo sapi
C 115	20.2	80.8	180126	14	AC145313	AC145313 Homo sapi
C 116	20.2	80.8	180970	14	AC140901	AC140901 Homo sapi
C 117	20.2	80.8	181842	14	AL391823	AL391823 Homo sapi
C 118	20.2	80.8	182374	5	BX649423	BX649423 Zebrafish
C 119	20.2	80.8	184155	9	AC124460	AC124460 Bos tauru
C 120	20.2	80.8	184254	14	AC166395	AC166395 Bos tauru
C 121	20.2	80.8	184872	9	AL672243	AL672243 Mouse DNA
C 122	20.2	80.8	186711	8	AC092144	AC092144 Homo sapi
C 123	20.2	80.8	186774	8	AC139887	AC139887 Homo sapi
C 124	20.2	80.8	187898	14	AC016715	AC016715 Homo sapi
C 125	20.2	80.8	187909	14	CR363422	CR363422 Danio rer
C 126	20.2	80.8	188845	14	CT025605	CT025605 Mus muscu
C 127	20.2	80.8	188887	9	AC122531	AC122531 Mus muscu
C 128	20.2	80.8	190237	9	AC153634	AC153634 Mus muscu
C 129	20.2	80.8	190430	14	AC118987	AC118987 Pan trogl
C 130	20.2	80.8	191405	14	AC139454	AC139454 Homo sapi
C 131	20.2	80.8	191853	9	AC116412	AC116412 Mus muscu
C 132	20.2	80.8	192118	14	AC140707	AC140707 Homo sapi
C 133	20.2	80.8	196359	14	AC142389	AC142389 Homo sapi
C 134	20.2	80.8	199028	14	AC128620	AC128620 Rattus no
C 135	20.2	80.8	199536	14	AC141067	AC141067 Homo sapi
C 136	20.2	80.8	200512	9	AC163106	AC163106 Mus muscu
C 137	20.2	80.8	201275	9	AL603662	AL603662 Mouse DNA
C 138	20.2	80.8	201302	9	AC154792	AC154792 Mus muscu
C 139	20.2	80.8	201808	14	AC141408	AC141408 Homo sapi
C 140	20.2	80.8	202655	9	AC161369	AC161369 Mus muscu
C 141	20.2	80.8	203870	8	AC140884	AC140884 Homo sapi
C 142	20.2	80.8	205924	9	AC155284	AC155284 Mus muscu
C 143	20.2	80.8	205933	14	BX927242	BX927242 Danio rer
C 144	20.2	80.8	206421	14	AC131110	AC131110 Homo sapi
C 145	20.2	80.8	211171	8	AC142381	AC142381 Homo sapi
C 146	20.2	80.8	212280	8	HS8362811	AL121873 Human DNA
C 147	20.2	80.8	213266	14	AC155752	AC155752 Bos tauru
C 148	20.2	80.8	213654	14	AC141609	AC141609 Homo sapi
C 149	20.2	80.8	214178	5	BX005375	BX005375 Zebrafish
C 150	20.2	80.8	214295	9	AC137947	AC137947 Mus muscu

ALIGNMENTS

RESULT 1
AC004738 147404 bp DNA linear PAT 25-NOV-1998
LOCUS Homo sapiens Chromosome 15q11-q13 PAC clone pDU351n23 from the
DEFINITION Prader-Willi/Angelman Syndrome region, complete sequence.
AC004738
AC004738.1 GI:3927853
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M., Buettner,J., Bumister,R., Card,P., desallboat,F., Dunn,J., English,C., Ehridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schegeman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
TITLE	HTGS Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 147404)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M., Buettner,J., Bumister,R., Card,P., desallboat,F., Dunn,J., English,C., Ehridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., McFarland,J., Miller,R., Newton,J., Osborne-Lawrence,S., Schegeman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
TITLE	Direct Submisison
JOURNAL	Submitted (23-MAY-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE	3 (bases 1 to 147404)
AUTHORS	Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M., Buettner,J., Butler,C., Card,P., desallboat,F., Dunn,J., English,C., Ehridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schegeman,J., Schultz,R.A., Stimson,S., Waller,K. and Ward,T.
TITLE	Direct Submisison
JOURNAL	Submitted (25-NOV-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT	On Nov 25, 1998, this sequence version replaced gi:315623. CHROMOSOMAL LOCUS: This clone comes from the Prader-Willi/Angelman Syndrome region mapped between STS D15S518 and D15S1019. MAPPED CLONE OVERLAP: PACs pDU14112 and pDU373b1. CONFIRMED MARKERS: sequence confirmed STS; D15S986.
FEATURES	Location/Qualifiers
source	1..147404
repeat_region	/organism="Homo sapiens"
repeat_region	/mol_type="Genomic DNA"
repeat_region	/db_xref="taxon:9606"
repeat_region	complement(1..38)
repeat_region	/rpt_family="Alu"
repeat_region	2480..2562
repeat_region	/rpt_family="MIR"
repeat_region	complement(7081..7180)
repeat_region	/rpt_family="Alu"
repeat_region	7753..8042
repeat_region	/rpt_family="Alu"
repeat_region	11636..12120
repeat_region	/rpt_family="MER1"
repeat_region	complement(12607..12704)
repeat_region	/rpt_family="Alu"
repeat_region	17422..17527
repeat_region	/rpt_family="Alu"
repeat_region	complement(17550..18923)
repeat_region	/rpt_family="L1"
repeat_region	complement(19315..20487)
repeat_region	/rpt_family="MER7"
repeat_region	complement(23589..23730)
repeat_region	/rpt_family="MIR"
repeat_region	26542..26840
repeat_region	/rpt_family="Alu"
repeat_region	complement(28089..28370)
repeat_region	/rpt_family="Alu"
repeat_region	30392..30756
repeat_region	/rpt_family="THE1"
repeat_region	33403..33491
repeat_region	/rpt_family="MER3"

```

repeat_region      33859. .34157
                    /rpt_family="Alu"
repeat_region      complement(34541. .34839)
                    /rpt_family="Alu"
repeat_region      complement(40268. .40564)
                    /rpt_family="L1"
repeat_region      40631. .40920
                    /rpt_family="Alu"
repeat_region      40925. .41391
                    /rpt_family="L1"
repeat_region      41483. .41761
                    /rpt_family="Alu"
repeat_region      complement(42533. .43025)
                    /rpt_family="Alu"
repeat_region      complement(43127. .43257)
                    /rpt_family="Alu"
repeat_region      46988. .47263
                    /rpt_family="Alu"
repeat_region      47905. .48181
                    /rpt_family="Alu"
repeat_region      49288. .49583
                    /rpt_family="Alu"
repeat_region      50805. .51226
                    /rpt_family="L1"
repeat_region      54069. .56429
                    /rpt_family="L1"
repeat_region      57563. .57629
                    /rpt_family="L1"
repeat_region      57873. .57916
                    /rpt_family="THE1"
repeat_region      complement(58196. .58265)
                    /rpt_family="THE1"
repeat_region      complement(58581. .58853)
                    /rpt_family="Alu"
repeat_region      complement(61377. .61968)
                    /rpt_family="LTR12"
repeat_region      69723. .69838
                    /rpt_family="MLT1"
repeat_region      70186. .70252
                    /rpt_family="MLT1"
repeat_region      70355. .70509
                    /rpt_family="MLT1"
repeat_region      73626. .73759
                    /rpt_family="L1"
repeat_region      complement(75899. .76175)
                    /rpt_family="Alu"
repeat_region      76595. .76779
                    /rpt_family="L1"
repeat_region      80584. .80830
                    /rpt_family="Alu"
repeat_region      81120. .81375
                    /rpt_family="L1"
repeat_region      81392. .81584
                    /rpt_family="L1"
repeat_region      81510. .82283
                    /rpt_family="MER42"
repeat_region      83319. .83585
                    /rpt_family="Alu"
repeat_region      complement(86469. .89229)
                    /rpt_family="L1"
repeat_region      complement(89216. .89402)
                    /rpt_family="Alu"
repeat_region      complement(92169. .92774)
                    /rpt_family="Alu"
repeat_region      complement(93392. .93583)
                    /rpt_family="MLT2B2"
repeat_region      complement(96114. .96232)
                    /rpt_family="MER7"
repeat_region      complement(96336. .96749)
                    /rpt_family="MLT1"
repeat_region      97775. .98173
                    /rpt_family="Alu"
repeat_region      complement(10095. .100225)

```

```

repeat_region      /rpt_family="MER"
100872. .101168
/rpt_family="Alu"
repeat_region      103147. .103264
/rpt_family="Alu"
repeat_region      10427. .104714
/rpt_family="Alu"
repeat_region      complement(105296. .105587)
/rpt_family="Alu"
repeat_region      complement(105678. .106117)
/rpt_family="L1"
repeat_region      complement(106162. .106459)
/rpt_family="Alu"
repeat_region      106728. .106810
/rpt_family="Alu"
repeat_region      108299. .108519
/rpt_family="L1"
repeat_region      109058. .109562
/rpt_family="L1"
repeat_region      109914. .110066
/rpt_family="Alu"
repeat_region      110321. .110606
/rpt_family="Alu"
repeat_region      111104. .111156
/rpt_family="MER42"
repeat_region      111716. .112012
/rpt_family="Alu"
repeat_region      112672. .112802
/rpt_family="Alu"
repeat_region      112823. .113204
/rpt_family="MLT1"
repeat_region      complement(113370. .113640)
/rpt_family="Alu"
repeat_region      119492. .119763
/rpt_family="MER2"
repeat_region      120000. .120270
/rpt_family="Alu"
repeat_region      complement(120965. .121242)
/rpt_family="Alu"
repeat_region      complement(121456. .121558)
/rpt_family="MLT1"
repeat_region      124613. .124768
/rpt_family="LTR9"
repeat_region      complement(125001. .125710)
/rpt_family="LTR12"
repeat_region      125759. .126009
/rpt_family="LTR9"

```

Query Match 100.0%; Score 25; DB 8; Length 147404;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTGATCTT 25
 Db 87804 AAAAAAAAACTAGCTGATCTT 87828

RESULT 2
 AC124305/c AC124305 28143 bp DNA linear PRI 06-FEB-2003
 LOCUS Homo sapiens chromosome 15, clone RP11-173H16, complete sequence.
 DEFINITION AC124305
 ACCESSION AC124305.3 GI:28261592
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.
 1 (bases 1 to 28143)
 AUTHORS Birren,B., Nuebaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 15, clone RP11-173H16
 JOURNAL Unpublished


```

repeat_region      8891..9182
                    /rpt_family="AluSx"
repeat_region      9183..9761
                    /rpt_family="L1ME2"
repeat_region      9762..10067
                    /rpt_family="AluSx"
repeat_region      10068..10315
                    /rpt_family="L1ME2"
repeat_region      10316..10374
                    /rpt_family="MIR"
repeat_region      10626..10933
                    /rpt_family="AluSx"
repeat_region      12081..12213
                    /rpt_family="FLAM_C"
repeat_region      12527..12562
                    /rpt_family="L2"
repeat_region      12771..12874
                    /rpt_family="A-rich"
repeat_region      14178..14485
                    /rpt_family="AluSg"
repeat_region      15119..15264
                    /rpt_family="MIR"
repeat_region      15640..16059
                    /rpt_family="LTR35"
repeat_region      16060..16105
                    /rpt_family="(TC)n"
repeat_region      16106..16217
                    /rpt_family="LTR35"
repeat_region      16223..17125
                    /rpt_family="L1ME3A"
repeat_region      17158..17461
                    /rpt_family="AluSx"
repeat_region      17465..17571
                    /rpt_family="AluSg/x"
repeat_region      17650..17734
                    /rpt_family="BLACKJACK"
repeat_region      17745..18059
                    /rpt_family="L1ME3A"
repeat_region      18065..18154
                    /rpt_family="MLT1D"

```

Query Match 89.64; Score 22.4; DB 8; Length 28143;
 Best Local Similarity 95.84; Pred. No. 2e+02; Mismatches 1; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 AAAAAAAAACTAGCTTGATCTT 25
 Db 27523 AAAAAAAAACTAGCTTGATCTT 27500

RESULT 3
 AC124311 142745 bp DNA linear PRI 08-JAN-2003
 LOCUS Homo sapiens chromosome 15, clone RP11-479F18, complete sequence.
 DEFINITION AC124311
 ACCESSION AC124311.7 GI:27545109
 VERSION HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 142745)
 Birren,B., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 15, clone RP11-479F18
 Unpublished
 2 (bases 1 to 142745)
 Birren,B., Linton,J., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 142745)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagoz,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connell,P., O'Neil,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

REFERENCE
 AUTHORS
 Submitted (16-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 142745)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagoz,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connell,P., O'Neil,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Rise,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 5 (bases 1 to 142745)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagoz,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connell,P., O'Neil,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Rise,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (125-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 5 (bases 1 to 142745)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagoz,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connell,P., O'Neil,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Rise,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.


```

repeat_region      24229..24321
                    /rpt_family="GA-rich"
repeat_region      complement(24880..24968)
                    /rpt_family="LIMAA4a"
repeat_region      25716..26029
                    /rpt_family="AluY"
repeat_region      26166..27037
                    /rpt_family="LIMB1"
repeat_region      27038..27341
                    /rpt_family="AluSg"
repeat_region      27342..27860
                    /rpt_family="LIMB1"
repeat_region      27861..28169
                    /rpt_family="AluSc"
repeat_region      28170..29250
                    /rpt_family="LIMB1"
repeat_region      29251..30314
                    /rpt_family="LIMC1"
repeat_region      30315..30750
                    /rpt_family="LIMC1"
repeat_region      complement(30765..30829)
                    /rpt_family="Tigres5"
repeat_region      30830..30923
                    /rpt_family="LIMC1"
repeat_region      30924..31228
                    /rpt_family="AluY"

Query Match      89.6%; Score 22.4; DB 8; Length 179810;
Best Local Similarity 95.8%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAAAAAAAACTAGCTTATCTT 25
DB      127747 AAAAAAAAACTAGCTTATCTT 172770

RESULT 5
AC165510
LOCUS      AC165510
DEFINITION      Bos taurus clone CH240-175C7, *** SEQUENCING IN PROGRESS ***, 16
unoriented pieces.
ACCESSION      AC165510
VERSION      AC165510.1 GI:70912528
KEYWORDS      HTG: HTGS_PHASE1.
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 51096)
Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P.,
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
Beraducci,A., Biswal,O., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
Cheng,M.-T., Chu,B., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H.,
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dzulda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flaegge,N., Forbes,L., Foster,P.,
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
Gonzalez-Garay,M., Guevara,W., Holder,M., Haeleand,W., Haeblerlein,K.,
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgson,A., Hogue,M., Hollins,B.,
Hoswell,L.T., Huliy,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafut,K., Kelly,S.,
Key,T., Khan,Z., King,L., Kovar,C., Kowals,A., Kowals,C., Lara,F.,
Leal,S., Lee,K., Lee,S., Legall,F.I., Lemmon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Llyanage,D.,
London,P., Lopez,J., Lorenshewe,L., Lozado,R., Luk,T., Madu,R.,

```

COMMENT

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help.tmc.edu

----- Project Information
Center project name: FKBA
Center clone name: CH240-175C7
----- Summary Statistics
Sequencing vector: plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 55091 bases at least Q40
Consensus quality: 58158 bases at least Q30
Consensus quality: 60554 bases at least Q20
Estimated insert size: 67341; sum-of-coverage estimation
Estimated insert size: 62018; agarose-fp estimation
Quality coverage: 1x in Q20 bases; agarose-fp estimation
Quality coverage: 1x in Q20 bases; sum-of-coverage estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 2283: contig of 2283 bp in length
* 2284: gap of unknown length
* 2384: contig of 2019 bp in length
* 4402: gap of unknown length
* 4403: gap of unknown length
* 4503: gap of 2517 bp in length
* 7020: gap of unknown length
* 7120: contig of 3991 bp in length
* 11110: gap of unknown length
* 11210: gap of 2773 bp in length
* 13983: gap of unknown length
* 14083: gap of 2034 bp in length
* 14084: gap of unknown length
* 16117: gap of unknown length
* 16218: gap of 2867 bp in length
* 19084: contig of 19184: gap of unknown length
* 19185: gap of 2760 bp in length
* 21945: gap of unknown length
* 22044: gap of unknown length
* 27322: contig of 5278 bp in length

```

```

*      27323      27422: gap of unknown length
*      27423      29952: contig of 2530 bp in length
*      29953      30052: gap of unknown length
*      30053      34263: contig of 4211 bp in length
*      34264      34363: gap of unknown length
*      34364      37425: contig of 3062 bp in length
*      37426      37525: gap of unknown length
*      37526      39785: contig of 2260 bp in length
*      39786      39885: gap of unknown length
*      39886      45612: contig of 5727 bp in length
*      45613      45712: gap of unknown length
*      45713      48745: contig of 3034 bp in length
*      48747      48846: gap of unknown length
*      48847      51096: contig of 2250 bp in length.
*
FEATURES
Source      1. .51096
            /organism="Bos taurus"
            /mol_type="genomic DNA"
            /db_xref="taxon:9913"
            /clone="CH240-175C7"
            2284. .2383
            /estimated_length=unknown
            4403. .4502
            /estimated_length=unknown
            7020. .7119
            /estimated_length=unknown
            1111. .11210
            /estimated_length=unknown
            13984. .14083
            /estimated_length=unknown
            16118. .16217
            /estimated_length=unknown
            19085. .19184
            /estimated_length=unknown
            21945. .22044
            /estimated_length=unknown
            27323. .27422
            /estimated_length=unknown
            29953. .30052
            /estimated_length=unknown
            34264. .34363
            /estimated_length=unknown
            37426. .37525
            /estimated_length=unknown
            39786. .39885
            /estimated_length=unknown
            45613. .45712
            /estimated_length=unknown
            48747. .48846
            /estimated_length=unknown

```

ORIGIN

```

Query Match      87.2%; Score 21.8; DB 14; Length 51096;
Best Local Similarity 92.0%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAACTATGCTTGCCTT 25
        ||||| ||||| ||||| |||||
Db      31019 AAAAAAAAAAGTATGCTTGCCTT 31043

```

```

RESULT 6
AC112769
LOCUS      AC112769      214779 bp      DNA      linear      HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-337C1, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION AC112769
VERSION AC112769.5 GI:25072673
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
          Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

```

REFERENCE

AUTHORS

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 214779)

Muzny,D,Marlee, Metzker,M, Lee, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Ayegagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cre, A., D'Souza, L.,

Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Davies, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebreygeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,

Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpethy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kows, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensheva, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,

Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Mulas, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwachukwu, O., Okunolu, G., Olamugbesun, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C.,

Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

River, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

Sanders, W., Savary, G., Scherer, S., Scott, G., Shetman, S., Shen, H.,

Shetty, J., Shvartbeyn, A., Sison, I., Sitter, C.D., Smajls, D.,

Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Soza, J.,

Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,

Valas, R., Vera, V., Villasana, D., Waldon, L., Walker, B., Wang, J.,

Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, U., Yoon, L., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von

Niederhausem, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,

Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSVZ
Center clone name: CH230-337C1

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 200558 bases at least Q40
Consensus quality: 202762 bases at least Q30
Consensus quality: 204578 bases at least Q20
Estimated insert size: 205128; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4464: contig of 4464 bp in length
* 4465 4564: gap of unknown length
* 4565 7854: contig of 3290 bp in length
* 7855 7954: gap of unknown length
* 7955 214779: contig of 206825 bp in length.
Location/Qualifiers

FEATURES
source
1..214779
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-337C1"

misc_feature
1..1270
/note="wgs end extension
clone_end:Sp6"
1987..4464
/note="wgs end extension
clone_end:Sp6"
4465..4564
/estimated_length=unknown
4565..5899
/note="wgs end extension
clone_end:Sp6"
7855..7954
/estimated_length=unknown
7955..11801
/note="wgs end extension
clone_end:Sp6"
8022..8851
/note="clone boundary
clone_end:Sp6"
site:
end_sequence:BN134521"
13478..15206
/note="wgs contig"
complement(203093..203995)
/note="clone boundary
clone_end:T7
site:
end_sequence:BN134520"

ORIGIN
Query Match

87.2%; Score 21.8; DB 14; Length 214779;

Best Local Similarity 92.0%; Pred. No. 1,9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTAGCTGATCTT 25
Db 159440 AAAAAAAAACTAGCTGATCTT 159464

RESULT 7
AC109007
LOCUS
DEFINITION
Rattus norvegicus clone CH230-42D14, *** SEQUENCING IN PROGRESS
*** 4 unordered pieces.
AC109007
AC109007.5 GI:24818592
VERSION
HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Behnmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Casar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cre, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Becotto, M., Eugene, C., Evans, C., Falle, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregregios, B., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, M.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, J., Kovar, C.,
Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozano, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M., McNeill, T., Meenan, B.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackemele, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., pfankoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primas, E., Fu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shetman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C., D., Smaj, D.,
Sneed, A., Sodergren, B., Song, X., Z., Sorelle, R., Sosa, D.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Woodson, H., Wolley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von
Weinhausen, A., Weiss, R., Smith, D., R., Holt, R., A., Smith, H., O.,
Weinstock, G., and Gibbs, R. A.

TITLE
JOURNAL
Unpublished

```

REFERENCE
AUTHORS      2 (bases 1 to 248019)
TITLE        Worley,K.C.
JOURNAL      Direct Submission
             Submitted (03-FEB-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3 (bases 1 to 248019)
AUTHORS      Rat Genome Sequencing Consortium.
TITLE        Direct Submission
             Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Nov 9, 2002 this sequence version replaced gi:23195541.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rac/). Each contig described
             in the feature table below represents a scaffold in the Atlas
             assembly ('a'='contig-scaffold'). Within each contig-scaffold,
             individual sequence contigs are ordered and oriented, and separated
             by sized gaps filled with Ns to the estimated size. The sequence
             may extend beyond the ends of the clone and there may be sequence
             contigs within a contig-scaffold that consist entirely of whole
             genome shotgun sequence reads. Both end sequences and whole genome
             shotgun sequence only contigs will be indicated in the feature
             table.
-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GPR
Center clone name: CH230-42D14
-----Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 205241 bases at least Q40
Consensus quality: 207774 bases at least Q30
Consensus quality: 209449 bases at least Q20
Estimated insert size: 212787; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_drafc\_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 169036: contig of 169036 bp in length
*   169037 169136: gap of unknown length
*   169137 245139: contig of 76003 bp in length
*   245140 245239: gap of unknown length
*   245240 246325: contig of 1086 bp in length
*   246326 246425: gap of unknown length
*   246426 248019: contig of 1594 bp in length.
Location/Qualifiers
1..248019
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-42D14"
18349..18927
/note="clone_boundary"
clone_end:T7
site:EcoRI
end_sequence:BH347216"
54431..54520
/note="clone_boundary"
clone_end:Spc
site:EcoRI
misc_feature
18349..18927
/misc_feature
54431..54520
/misc_feature
site:EcoRI
end_sequence:BH347216"
-----
gap          end_sequence:BH347216"
              169037..169136
              /estimated_length=unknown
misc_feature 169137..171084
              /note="wgs_end_extension"
              clone_end:Spc"
              171375..174392
              /note="wgs_end_extension"
              clone_end:Spc"
              245140..245239
              /estimated_length=unknown
              246326..246425
              /estimated_length=unknown
ORIGIN
Query Match      87.2%; Score 21.8; DB 14; Length 248019;
Best Local Similarity 92.0%; Pred No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy               1 AAAAAAAAAAAGTATGCTGATCTT 25
Db               146150 AAAAAAAAAAAGTATGCTGATCTT 146174
-----
RESULT 8
AC072057/c     140150 bp   DNA       linear   PRI 07-NOV-2001
LOCUS          Homo sapiens BAC clone RP11-312C1 from 7, complete sequence.
DEFINITION
AC072057
VERSION        AC072057.8 GI:16756346
KEYWORDS
SOURCE         HTG.
ORGANISM       Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homimidae; Homo.
1 (bases 1 to 140150)
Suleston,J.E. and Waterston,R.
Toward a complete human genome sequence
JOURNAL
PUBMED        9847074
REFERENCE      2 (bases 1 to 140150)
AUTHORS       Heyen,J., Haakenson,W., Hawkins,M. and Rose,C.
TITLE         The sequence of Homo sapiens BAC clone RP11-312C1
JOURNAL
REFERENCE      3 (bases 1 to 140150)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
SUBMITTED     (07-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 140150)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 7, 2001 this sequence version replaced gi:13431138.
-----Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplen@watscn.wustl.edu
-----Summary Statistics
Center project name: H_NH0312C01
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate

```

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenou, M., Catanese, V.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-876K2, 200 bp overlap. Actual start of this clone is at base position 94341 of RP5-876K2; actual end is at base position 140150 of RP11-312C1.

Location/Qualifiers

FEATURES

source

```
1..140150
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7"
  /clone="RP11-312C1"
  /clone_11b="RPCI-11"
  /rpt_family="MaLR"
  repeat_region
    385..747
  /rpt_family="MaLR"
  repeat_region
    1223..1592
  /rpt_family="MaLR"
  repeat_region
    1906..2337
  /rpt_family="ERV1"
  repeat_region
    3490..3800
  /rpt_family="ERV1"
  repeat_region
    3801..3853
  /rpt_family="(TTAGGG)n"
  repeat_region
    3854..3910
  /rpt_family="ERV1"
  repeat_region
    6369..6674
  /rpt_family="Alu"
  /rpt_family="Alu"
  8115..8167
  /note="similar to EST BB031171 (NTD:98326180)"
  repeat_region
    8126..8435
  /rpt_family="Alu"
  repeat_region
    8497..8692
  /rpt_family="MERS3"
  repeat_region
    9283..9305
  /rpt_family="(TTC)n"
  misc_feature
    9399..9878
  /note="similar to EST AA570380 (NTD:92344360) nK62D05.81"
  repeat_region
    9318..9437
  /rpt_family="MIR"
  repeat_region
    9738..10130
  /rpt_family="MaLR"
  repeat_region
    10682..10776
  /rpt_family="MIR"
  repeat_region
    10777..11083
  /rpt_family="Alu"
  repeat_region
    11084..11164
  /rpt_family="MIR"
```

```
repeat_region 11249..11522
  /rpt_family="MaLR"
repeat_region 11634..11695
  /rpt_family="MIR1_type"
repeat_region 11696..11996
  /rpt_family="Alu"
repeat_region 12014..12349
  /rpt_family="L1"
repeat_region 12723..12835
  /rpt_family="CR1"
repeat_region 13075..13125
  /rpt_family="(CA)n"
repeat_region 13178..13473
  /rpt_family="Alu"
repeat_region 13541..13650
  /rpt_family="L1"
repeat_region 14292..14466
  /rpt_family="ERV1"
repeat_region 14467..14745
  /rpt_family="Alu"
repeat_region 14746..14989
  /rpt_family="ERV1"
repeat_region 14990..15289
  /rpt_family="Alu"
repeat_region 15294..15371
  /rpt_family="(TA)n"
repeat_region 15373..15517
  /rpt_family="L1"
repeat_region 15549..15672
  /rpt_family="ERV1"
misc_feature 16703..16950
  /note="similar to EST AW841983 (NTD:97935966)"
repeat_region 16832..17199
  /rpt_family="ERV1"
repeat_region 17982..18134
  /rpt_family="MIR1_type"
repeat_region 18326..18617
  /rpt_family="Alu"
repeat_region 18769..19226
  /rpt_family="MaLR"
repeat_region 19646..20057
  /rpt_family="MaLR"
repeat_region 20058..20343
  /rpt_family="Alu"
repeat_region 20347..20459
  /rpt_family="L1"
repeat_region 20460..20478
  /rpt_family="Alu"
repeat_region 22638..22767
  /rpt_family="Alu"
repeat_region 22768..22830
  /rpt_family="L1"
repeat_region 22840..23010
  /rpt_family="L1"
repeat_region 23035..23345
  /rpt_family="Alu"
repeat_region 23542..23837
  /rpt_family="Alu"
repeat_region 25415..25540
  /rpt_family="L1"
repeat_region 25541..27466
  /rpt_family="L1"
repeat_region 27467..27823
  /rpt_family="L1"
misc_feature 27851..28304
  /note="similar to EST BF890766 (NTD:91282225)"
repeat_region 30130..30195
  /rpt_family="MaLR"
repeat_region 31277..31471
  /rpt_family="ERV1"
repeat_region 32060..32210
  /rpt_family="AchoBo"
repeat_region 32907..33339
```



```

repeat_region      /rpt_family="MALKR"
                   33435..33499
repeat_region      /rpt_family="Alu"
                   33500..33784
repeat_region      /rpt_family="Alu"
                   33785..33807
repeat_region      /rpt_family="AT_rich"

```

```

Query Match      85.6%; Score 21.4; DB 8; Length 140150;
Best Local Similarity 95.7%; Pred. No. 2.9e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      3 AAAAAAAAACTATGCTGATCTT 25
        |||||
Db      40953 AAAAAAAAACTATGCTGATCTT 40931

```

```

RESULT 9
BV456584/c      628 bp      DNA      linear      STS 06-APR-2005
LOCUS      gf79604.b1 Clint Pan troglodytes versus STS genomic, sequence
DEFINITION      tagged site.
ACCESSION      BV456584
VERSION      BV456584.1 GI:62216012
KEYWORDS      STS:
SOURCE      Pan troglodytes versus
ORGANISM      Pan troglodytes versus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Pan.
1 (bases 1 to 628)
Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
Jaffe,D.B.
Initial Sequence of the Chimpanzee Genome and Comparison with the
Human Genome
Unpublished (2005)

```

JOURNAL COMMENT

Contact: Michael C. Zody
Broad Institute of MIT and Harvard
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172580933
Fax: 6172580903
Email: mczody@broad.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 628

Protocol:
23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes versus chimps (Donald, Karlien, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes versus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have phred score >= 20, at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has phred score >= 30, the surrounding 10 bases in the read have phred score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less

than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

FEATURES

```

source
1..628
/organism="Pan troglodytes versus"
/mol_type="genomic DNA"
/sub_species="versus"
/db_xref="taxon:37012"
/clone_lib="Clint"
<1..>628

```

ORIGIN

```

Query Match      83.2%; Score 20.8; DB 10; Length 628;
Best Local Similarity 91.7%; Pred. No. 2.2e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      2 AAAAAAAAACTATGCTGATCTT 25
        |||||
Db      532 AAAAAAAAACTATGCTGATCTT 509

```

```

RESULT 10
BV491883      724 bp      DNA      linear      STS 07-APR-2005
LOCUS      S221P61080RH11.T0 Yvonne Pan troglodytes troglodytes STS genomic,
DEFINITION      sequence tagged site.
ACCESSION      BV491883
VERSION      BV491883.1 GI:62335553
KEYWORDS      STS:
SOURCE      Pan troglodytes troglodytes
ORGANISM      Pan troglodytes troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Pan.
1 (bases 1 to 724)
Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
Jaffe,D.B.
Initial Sequence of the Chimpanzee Genome and Comparison with the
Human Genome
Unpublished (2005)

```

JOURNAL COMMENT

Contact: Michael C. Zody
Broad Institute of MIT and Harvard
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172580933
Fax: 6172580903
Email: mczody@broad.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 724

Protocol:
23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes versus chimps (Donald, Karlien, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes versus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have phred score >= 20, at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has phred score >= 30, the surrounding 10 bases in


```

gap          gap        9462..9561      /estimated_length=unknown
gap          gap        12309..12508    /estimated_length=unknown
gap          gap        15277..15376    /estimated_length=unknown
gap          gap        18572..18671    /estimated_length=unknown
gap          gap        20238..20337    /estimated_length=unknown
gap          gap        22638..22737    /estimated_length=unknown
gap          gap        26523..26622    /estimated_length=unknown
gap          gap        31630..31729    /estimated_length=unknown
gap          gap        37688..37787    /estimated_length=unknown
gap          gap        43983..44082    /estimated_length=unknown
gap          gap        49534..49633    /estimated_length=unknown
gap          gap        57758..57857    /estimated_length=unknown
gap          gap        67594..67693    /estimated_length=unknown
gap          gap        77746..77845    /estimated_length=unknown
gap          gap        /estimated_length=unknown

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 87782;
Best Local Similarity 91.7%; Pred. No. 5.5e+02;
Matches 22; Conservative 0; Indels 0; Gaps 0;

Cy       2 AAAAAAAAACTAGTGTGATCTT 25
         |||||
Db       79778 AAAAAAAAACTTAGATTGATCTT 79601

RESULT 14
LOCUS     ALJ58954/C              88311 bp   DNA           linear   HTG_13-JUN-2001
DEFINITION Homo sapiens chromosome 6 clone RP3-33805, 2 unordered pieces.
ACCESSION ALJ58954
VERSION   ALJ58954.9 GI:10186568
KEYWORDS  HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
1
Sims,S.
Direct Submission
Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requester: clonerquest@sanger.ac.uk
On Sep 19, 2000 this sequence version replaced gi:10129499.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: d033805
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; l08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 88097 bases at least Q40
Consensus quality: 88159 bases at least Q30
Consensus quality: 88188 bases at least Q20
Insert size: 88211; sum-of-conflicts

```

```

Insert size: 155031; 40.3% error; agarose-gel
Quality coverage: 6.97% in Q20 bases; sum-of-contigs Quality
coverage: 4.03% in Q20 bases; agarose-gel
-----
** NOTE: This is a 'working draft' sequence. It currently
** consists of 2 contigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the contigs are represented as
** runs of N, but the exact sizes of the gaps are unknown.
** This record will be updated with the finished sequence
** as soon as it is available and the accession number will
** be preserved.
*      1      22000: contig of 22000 bp in length
*      *      22001      22100: gap of 100 bp
*      *      22101      88311: contig of 66211 bp in length.
FEATURES
    Source
        Location/Qualifiers
            1..88311
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="6"
                /clone_id="RP3-33805"
                /clone_1kb="RPCT-3"
                1..22000
                    /note="assembly_fragment:01291
                        clone_end:SP6
                        vector_side:left"
                        22101..88311
                            /note="assembly_fragment:00403
                                clone_end:T7
                                vector_side:right"
ORIGIN
Query Match          83.2%; Score 20.8; DB 14; Length 88311;
Best Local Similarity 91.7%; Pred. No. 5.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy      1 AAAAAAAAAAATAGCTTGATCT 24
         |||||
Db       25302 AAAAAAAAGCTTAAGCTTGATCT 25279
RESULT 15
AP007366/c
LOCUS
DEFINITION
Accession
Version
Keywords
Source
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT
1 Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
Structural Analysis of a Lotus japonicus Genome. XI. Sequence
Features and Mapping of Nine hundred twenty-one TAC Clones
Unpublished
2 (bases 1 to 96008)
Sato,S.
Direct Submission
Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu,
Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp),
URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(fax.2337),
Fax:81-438-52-3934)
** NOTE: This is a 'working draft' sequence. It currently
** consists of 26 contigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the contigs are represented as

```

```
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1830: contig of 1830 bp in length
*      1831      1930: gap of unknown length
*      1931      2990: contig of 1060 bp in length
*      2991      3090: gap of unknown length
*      3091      4056: contig of 966 bp in length
*      4057      4156: gap of unknown length
*      4157      5150: contig of 994 bp in length
*      5151      5250: gap of unknown length
*      5251      6043: contig of 793 bp in length
*      6044      6143: gap of unknown length
*      6144      7023: contig of 880 bp in length
*      7024      7123: gap of unknown length
*      7124      8236: contig of 1113 bp in length
*      8237      8336: gap of unknown length
*      8337      9615: contig of 1279 bp in length
*      9616      9715: gap of unknown length
*      9716      10938: contig of 1223 bp in length
*      10939      11038: gap of unknown length
*      11039      12497: contig of 1459 bp in length
*      12498      12597: gap of unknown length
*      12598      13931: contig of 1334 bp in length
*      13932      14031: gap of unknown length
*      14032      16464: contig of 2433 bp in length
*      16465      16564: gap of unknown length
*      16565      18712: contig of 2148 bp in length
*      18713      18812: gap of unknown length
*      18813      20468: contig of 1656 bp in length
*      20469      20568: gap of unknown length
*      20569      23047: contig of 2479 bp in length
*      23048      23147: gap of unknown length
*      23148      26681: contig of 3534 bp in length
*      26682      26781: gap of unknown length
*      26782      31099: contig of 4318 bp in length
*      31100      31199: gap of unknown length
*      31200      36348: contig of 5149 bp in length
*      36349      36448: gap of unknown length
*      36449      41923: contig of 5475 bp in length
*      41924      42023: gap of unknown length
*      42024      48370: contig of 6347 bp in length
*      48371      48470: gap of unknown length
*      48471      53746: contig of 5276 bp in length
*      53747      53846: gap of unknown length
*      53847      58556: contig of 4710 bp in length
*      58557      58656: gap of unknown length
*      58657      67011: contig of 8355 bp in length
*      67012      67111: gap of unknown length
*      67112      79048: contig of 11937 bp in length
*      79049      79148: gap of unknown length
*      79149      90009: contig of 10861 bp in length
*      90010      90109: gap of unknown length
*      90110      96008: contig of 5899 bp in length.
*
Location/Qualifiers
1. 96008
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="takon:34305"
/clone="LJT12L18"
/clone_lib="LJT library"
/note="TAC clone:TW0457, synonym:Lotus japonicus"
1831. 1930
/estimated_length=unknown
2991. 3090
/estimated_length=unknown
4057. 4156
/estimated_length=unknown
5151. 5250
/estimated_length=unknown
6044. 6143
```

```
gap      /estimated_length=unknown
gap      7024. 7123
gap      /estimated_length=unknown
gap      8237. 8336
gap      /estimated_length=unknown
gap      9616. 9715
gap      /estimated_length=unknown
gap      10939. 11038
gap      /estimated_length=unknown
gap      12498. 12597
gap      /estimated_length=unknown
gap      13932. 14031
gap      /estimated_length=unknown
gap      16465. 16564
gap      /estimated_length=unknown
gap      18713. 18812
gap      /estimated_length=unknown
gap      20469. 20568
gap      /estimated_length=unknown
gap      23048. 23147
gap      /estimated_length=unknown
gap      26682. 26781
gap      /estimated_length=unknown
gap      31100. 31199
gap      /estimated_length=unknown
gap      36349. 36448
gap      /estimated_length=unknown
gap      41924. 42023
gap      /estimated_length=unknown
gap      48371. 48470
gap      /estimated_length=unknown
gap      53747. 53846
gap      /estimated_length=unknown
gap      58557. 58656
gap      /estimated_length=unknown
gap      67012. 67111
gap      /estimated_length=unknown
gap      79049. 79148
gap      /estimated_length=unknown
gap      90010. 90109
gap      /estimated_length=unknown

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 96008;
Best Local Similarity 91.7%; Pred. No. 5.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAAAAAAAACTAGCTGATCTT 25
DB      92046 AAAAAAAAAGTAGATGATCTT 92023

RESULT 16
AP008208_146/c
WPCOMMENT
Sequence split into 360 fragments LOCUS AP008208 Accession AP008208
Fragment Name      Begin      End
AP008208_000      1      110000
AP008208_001      200001      210000
AP008208_002      200001      310000
AP008208_003      300001      410000
AP008208_004      400001      510000
AP008208_005      500001      610000
AP008208_006      600001      710000
AP008208_007      700001      810000
AP008208_008      800001      910000
AP008208_009      900001      1010000
AP008208_010      1000001      1110000
AP008208_011      1100001      1210000
AP008208_012      1200001      1310000
AP008208_013      1300001      1410000
AP008208_014      1400001      1510000
AP008208_015      1500001      1610000
```

AP008208_016 1600001 1710000
AP008208_017 1700001 1810000
AP008208_018 1800001 1910000
AP008208_019 1900001 2010000
AP008208_020 2000001 2110000
AP008208_021 2100001 2210000
AP008208_022 2200001 2310000
AP008208_023 2300001 2410000
AP008208_024 2400001 2510000
AP008208_025 2500001 2610000
AP008208_026 2600001 2710000
AP008208_027 2700001 2810000
AP008208_028 2800001 2910000
AP008208_029 2900001 3010000
AP008208_030 3000001 3110000
AP008208_031 3100001 3210000
AP008208_032 3200001 3310000
AP008208_033 3300001 3410000
AP008208_034 3400001 3510000
AP008208_035 3500001 3610000
AP008208_036 3600001 3710000
AP008208_037 3700001 3810000
AP008208_038 3800001 3910000
AP008208_039 3900001 4010000
AP008208_040 4000001 4110000
AP008208_041 4100001 4210000
AP008208_042 4200001 4310000
AP008208_043 4300001 4410000
AP008208_044 4400001 4510000
AP008208_045 4500001 4610000
AP008208_046 4600001 4710000
AP008208_047 4700001 4810000
AP008208_048 4800001 4910000
AP008208_049 4900001 5010000
AP008208_050 5000001 5110000
AP008208_051 5100001 5210000
AP008208_052 5200001 5310000
AP008208_053 5300001 5410000
AP008208_054 5400001 5510000
AP008208_055 5500001 5610000
AP008208_056 5600001 5710000
AP008208_057 5700001 5810000
AP008208_058 5800001 5910000
AP008208_059 5900001 6010000
AP008208_060 6000001 6110000
AP008208_061 6100001 6210000
AP008208_062 6200001 6310000
AP008208_063 6300001 6410000
AP008208_064 6400001 6510000
AP008208_065 6500001 6610000
AP008208_066 6600001 6710000
AP008208_067 6700001 6810000
AP008208_068 6800001 6910000
AP008208_069 6900001 7010000
AP008208_070 7000001 7110000
AP008208_071 7100001 7210000
AP008208_072 7200001 7310000
AP008208_073 7300001 7410000
AP008208_074 7400001 7510000
AP008208_075 7500001 7610000
AP008208_076 7600001 7710000
AP008208_077 7700001 7810000
AP008208_078 7800001 7910000
AP008208_079 7900001 8010000
AP008208_080 8000001 8110000
AP008208_081 8100001 8210000
AP008208_082 8200001 8310000
AP008208_083 8300001 8410000
AP008208_084 8400001 8510000
AP008208_085 8500001 8610000
AP008208_086 8600001 8710000
AP008208_087 8700001 8810000
AP008208_088 8800001 8910000

AP008208_089 8900001 9010000
AP008208_090 9000001 9110000
AP008208_091 9100001 9210000
AP008208_092 9200001 9310000
AP008208_093 9300001 9410000
AP008208_094 9400001 9510000
AP008208_095 9500001 9610000
AP008208_096 9600001 9710000
AP008208_097 9700001 9810000
AP008208_098 9800001 9910000
AP008208_099 9900001 10010000
AP008208_100 10000001 10110000
AP008208_101 10100001 10210000
AP008208_102 10200001 10310000
AP008208_103 10300001 10410000
AP008208_104 10400001 10510000
AP008208_105 10500001 10610000
AP008208_106 10600001 10710000
AP008208_107 10700001 10810000
AP008208_108 10800001 10910000
AP008208_109 10900001 11010000
AP008208_110 11000001 11110000
AP008208_111 11100001 11210000
AP008208_112 11200001 11310000
AP008208_113 11300001 11410000
AP008208_114 11400001 11510000
AP008208_115 11500001 11610000
AP008208_116 11600001 11710000
AP008208_117 11700001 11810000
AP008208_118 11800001 11910000
AP008208_119 11900001 12010000
AP008208_120 12000001 12110000
AP008208_121 12100001 12210000
AP008208_122 12200001 12310000
AP008208_123 12300001 12410000
AP008208_124 12400001 12510000
AP008208_125 12500001 12610000
AP008208_126 12600001 12710000
AP008208_127 12700001 12810000
AP008208_128 12800001 12910000
AP008208_129 12900001 13010000
AP008208_130 13000001 13110000
AP008208_131 13100001 13210000
AP008208_132 13200001 13310000
AP008208_133 13300001 13410000
AP008208_134 13400001 13510000
AP008208_135 13500001 13610000
AP008208_136 13600001 13710000
AP008208_137 13700001 13810000
AP008208_138 13800001 13910000
AP008208_139 13900001 14010000
AP008208_140 14000001 14110000
AP008208_141 14100001 14210000
AP008208_142 14200001 14310000
AP008208_143 14300001 14410000
AP008208_144 14400001 14510000
AP008208_145 14500001 14610000
AP008208_146 14600001 14710000
AP008208_147 14700001 14810000
AP008208_148 14800001 14910000
AP008208_149 14900001 15010000
AP008208_150 15000001 15110000
AP008208_151 15100001 15210000
AP008208_152 15200001 15310000
AP008208_153 15300001 15410000
AP008208_154 15400001 15510000
AP008208_155 15500001 15610000
AP008208_156 15600001 15710000
AP008208_157 15700001 15810000
AP008208_158 15800001 15910000
AP008208_159 15900001 16010000
AP008208_160 16000001 16110000
AP008208_161 16100001 16210000

AP008208_162 16200001 16310000
 AP008208_163 16300001 16410000
 AP008208_164 16400001 16510000
 AP008208_165 16500001 16610000
 AP008208_166 16600001 16710000
 AP008208_167 16700001 16810000
 AP008208_168 16800001 16910000
 AP008208_169 16900001 17010000
 AP008208_170 17000001 17110000
 AP008208_171 17100001 17210000
 AP008208_172 17200001 17310000
 AP008208_173 17300001 17410000
 AP008208_174 17400001 17510000
 AP008208_175 17500001 17610000
 AP008208_176 17600001 17710000
 AP008208_177 17700001 17810000
 AP008208_178 17800001 17910000
 AP008208_179 17900001 18010000
 AP008208_180 18000001 18110000
 AP008208_181 18100001 18210000
 AP008208_182 18200001 18310000
 AP008208_183 18300001 18410000
 AP008208_184 18400001 18510000
 AP008208_185 18500001 18610000
 AP008208_186 18600001 18710000
 AP008208_187 18700001 18810000
 AP008208_188 18800001 18910000
 AP008208_189 18900001 19010000
 AP008208_190 19000001 19110000
 AP008208_191 19100001 19210000
 AP008208_192 19200001 19310000
 AP008208_193 19300001 19410000
 AP008208_194 19400001 19510000
 AP008208_195 19500001 19610000
 AP008208_196 19600001 19710000

Query Match 83.2%; Score 20.8; DB 15; Length 110000;
 Best Local Similarity 91.7%; Pred. No. 5.2e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATGCTGATCT 24
 Db 40772 AAAAAAAAACTATGCTGAT 40749

RESULT 17
 AC093811 115335 bp DNA linear PRI 01-MAR-2002
 LOCUS AC093811
 DEFINITION Homo sapiens BAC clone RP11-35SH11 from 4, complete sequence.
 AC093811 AC024662
 VERSION AC093811.3 GI:16903168
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.
 1 (bases 1 to 115335)
 REFERENCE AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 PUBMED 9847074
 2 (bases 1 to 115335)
 REFERENCE AUTHORS Shah, N. and Haakenson, W.
 TITLE The sequence of Homo sapiens BAC clone RP11-35SH11
 JOURNAL Unpublished (2001)
 3 (bases 1 to 115335)
 REFERENCE AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 115335)
 REFERENCE

AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 115335)
 REFERENCE AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 6 (bases 1 to 115335)
 REFERENCE AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 11, 2001 this sequence version replaced gi:15624965.
 COMMENT ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH035SH11
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, B., Tateno, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the right is RP11-240A16, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-35SH11; actual end is at base position 80018 of RP11-240A16.

There is an unresolved base at 67647.

The sequence of AC024662 has been incorporated into AC093811.

FEATURES
 source
 Location/Qualifiers

1..115335
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-35SH11"
 /clone_11b="RP11-11"

```

repeat_region      213. 710
                    /rpt_family="BRVL"
repeat_region      880. 1022
                    /rpt_family="L2"
repeat_region      1315. 1568
                    /rpt_family="MALR"
repeat_region      1587. 1606
                    /rpt_family="MIR"
repeat_region      2357. 2377
                    /rpt_family="(TTAA)n"
repeat_region      4322. 4466
                    /rpt_family="MIR"
repeat_region      4487. 4630
                    /rpt_family="MALR"
repeat_region      5402. 5425
                    /rpt_family="(TTTG)n"
repeat_region      5460. 5487
                    /rpt_family="AT_rich"
repeat_region      5514. 5563
                    /rpt_family="(TA)n"
repeat_region      7007. 7144
                    /rpt_family="MIR"
repeat_region      7156. 7254
                    /rpt_family="L2"
repeat_region      7764. 8243
                    /rpt_family="L2"
repeat_region      7908. 7929
                    /rpt_family="AT_rich"
repeat_region      8263. 8637
                    /rpt_family="MALR"
repeat_region      11252. 11289
                    /rpt_family="(T)n"
repeat_region      11265. 11553
                    /rpt_family="Alu"
repeat_region      12887. 12909
                    /rpt_family="(TG)n"
repeat_region      13341. 13800
                    /rpt_family="ERVX"
repeat_region      14357. 15693
                    /rpt_family="L1"
repeat_region      15694. 15783
                    /rpt_family="L1"
repeat_region      15784. 16961
                    /rpt_family="L1"
repeat_region      16941. 16962
                    /rpt_family="AT_rich"
repeat_region      16962. 17261
                    /rpt_family="MER2_type"
repeat_region      17092. 17121
                    /rpt_family="AT_rich"
repeat_region      17893. 18071
                    /note="match to EST AV690508 (NID:gi0292371)"
misc_feature
repeat_region      18617. 18644
                    /rpt_family="AT_rich"
repeat_region      19215. 19261
                    /rpt_family="MIR"
repeat_region      20566. 20648
                    /rpt_family="(TA)n"
repeat_region      20650. 20670
                    /rpt_family="AT_rich"
repeat_region      21822. 21909
                    /rpt_family="ERV1"
repeat_region      22536. 22735
                    /rpt_family="L2"
repeat_region      22890. 23184
                    /rpt_family="Alu"
repeat_region      23767. 23802
                    /rpt_family="AT_rich"
repeat_region      23903. 24193
                    /rpt_family="Alu"
repeat_region      24195. 24220
                    /rpt_family="AT_rich"
repeat_region      24345. 24535

```

```

repeat_region      /rpt_family="L2"
                    25334. 25460
repeat_region      /rpt_family="ERV1"
                    25346. 25483
repeat_region      /rpt_family="(TATATG)n"
                    25782. 25891
repeat_region      /rpt_family="MER1_type"
                    25955. 26338
repeat_region      /rpt_family="MALR"
                    26407. 26440
repeat_region      /rpt_family="(T)n"
                    26411. 26704
repeat_region      /rpt_family="Alu"
                    26709. 26758
repeat_region      /rpt_family="Mariner"
                    27106. 27147
repeat_region      /rpt_family="(TTTA)n"
                    27138. 27269
repeat_region      /rpt_family="Alu"
                    27471. 27546
repeat_region      /rpt_family="MIR"
                    27540. 27576
repeat_region      /rpt_family="(CAT)n"
                    27980. 28004
repeat_region      /rpt_family="AT_rich"
                    27992. 28212
repeat_region      /rpt_family="L1"
                    28213. 28517

Query Match      83.2%; Score 20.8; DB 8; Length 11535;
Best Local Similarity 91.7%; Pred. No. 5.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTATAGCTGATCT 24
Db      113426 AAAAAAAAAAGCTTAAGCTTGATCT 113449

RESULT 18
AP004855/c      132282 bp DNA linear PLN 13-JUL-2004
LOCUS      Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
DEFINITION      BAC clone:OU1538_H05.
ACCESSION      AP004855
VERSION      AP004855.3 GI:50252108
KEYWORDS
ORGANISM      Oryza sativa (japonica cultivar-group)
SOURCE      Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Brizarioidae; Oryzaceae; Oryza.
REFERENCE
AUTHORS      Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE      Oryza sativa nipponbare (GAS) genomic DNA, chromosome 2, BAC
            clone:OU1538_H05
JOURNAL      Published Only in Database (2002)
AUTHORS      Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE      Direct Submission
JOURNAL      Submitted (20-MAR-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program, Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT      On Jul 12, 2004 this sequence version replaced gi:47522601.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark (
            http://opal.biology.gatech.edu/GeneMark/), GLIMMER
            (http://www.cigr.org/cdb/glimmer/glimr_form.html), RICEHMM
            (http://rtp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/gp.cgi), slm4
            (http://globin.cse.psu.edu/html/docs/slm4.html), gap2

```


(<http://www.tigr.org/software/glimmer/>), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database at RGP or DBJ. Protein homologues of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of OJ1538_H05 clone has an overlap with OSJNB0075E08 clone (DBJ: AP005653) at 5' end and an overlap with OSJNB0055113 (DBJ: AP005648) at 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomesSeq.html>.

FEATURES

SOURCE

```
1. 132282
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="OJ1538_H05"
/complement(join(58..125,260..454,799..831,1420..1477))
/gene="OJ1538_H05.1"
/feature(join(58..125,260..454,799..831,1420..1477))
/note="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
join(3357..3366,3472..3653)
/gene="OJ1538_H05.2"
join(3357..3366,3472..3653)
/note="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
4771..5004
/gene="OJ1538_H05.3"
4771..5004
/gene="OJ1538_H05.3"
/note="hypothetical ORF
predicted by FGENESH
this category is not included in IRGSP standard"
/gene="OJ1538_H05.4"
/complement(5767..6780)
/gene="OJ1538_H05.4"
/complement(5767..6780)
/note="3' terminal repeat"
/rpt_type="terminal"
/complement(8097..13413)
/gene="OJ1538_H05.4"
/note="probably inactive due to including stop codon(s) in
CDS
pseudogene, gag-pol polypeptide"
/pseudogene(join(13497..14515))
/gene="OJ1538_H05.4"
/note="5' terminal repeat"
/rpt_type="terminal"
```

```
16896..18231
/gene="OJ1538_H05.5"
join(<16896..16958,17429..17541,17841..>18231)
/gene="OJ1538_H05.5"
/note="start and end point are not identified"
join(16896..16958,17429..17541,17841..18231)
/gene="OJ1538_H05.5"
/note="predicted by FGENESH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD28094.1"
/db_xref="GI:50252109"
/translation="MSAIKTLTFREADLLVSGRMTRVTEGDTKIIIVEGDEIS
RKGVSVDGSGGSGGASQRRRRRGGEWIRHLLLPSPGDGIEGRWLAH
AAVLERKSIDGPPVPPPLGSGAGRAAGVALVNSRLSGTGRAGADPAP
PPSPPAAKRRDGIKVASAGGYL"
18654..18925
/gene="OJ1538_H05.6"
join(<18654..18742,18856..>18925)
/gene="OJ1538_H05.6"
/note="start and end point are not identified"
join(18654..18742,18856..18925)
/gene="OJ1538_H05.6"
/note="predicted by GlimmerM etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD28095.1"
/db_xref="GI:50252110"
/translation="MGSPDDENGSPWVKLVLAGSGDGLASGHPATEPVYQLDELVK
VWKKNLF"
complement(20261..27983)
/gene="OJ1538_H05.7"
/complement(join(20261..20602,21917..22045,22181..22534,
22615..22862,22953..23029,23372..23446,23519..23620,
23764..23882,26223..26346,26426..26476,26591..26689,
27760..27983))
/gene="OJ1538_H05.7"
/note="supported by full-length cDNA(s): AK064509"
complement(join(20474..20602,21917..22045,22181..22534,
22615..22862,22953..23029,23372..23446,23519..23620,
23764..23882,26223..26346,26426..26476,26591..26689,
27760..27983))
/gene="OJ1538_H05.7"
/note="Ca2+-dependent lipid-binding protein
contains EST(s): AU100617(C10221),D22040(C10221)
contains full-length cDNA(s): AK064509"
/codon_start=1
/product="putative CLB1 protein"
/protein_id="BAD28096.1"
/db_xref="GI:50252111"
/translation="MGFRTVLGFGFGVGTWGLVGYLFIYPOPTDVQDPVIRPL
VELDTKSLSEMLPEVPHWVKNPDRIDMLNKVENIWPFLDKAIKTAKEIKPVIA
ENAKYKIDSVEFETLTGLSPTPGMGVYTTDBEGLIMEPSIKKAGNPNTVWKA
FGIKATRAVYIDLVFALPRITLAKPLPSPPCPRAKIYVSLMEKPHVDPLKLCADLMA
IRLIFYFVPIITKTQVANNYLMFKVLVPIIMDPARQKQKQVGLHVNIVRAVLAITKDD
FLGKSDPYVKLTETKELPSKTSVRSKMLNPEWDFKLNVVDPSQALLETLYVYWE
OVGKHDKIMSVYPLKELIPDEAKSLTLDLHKTMNDANDPANDFRQLTVDYVKEPK
EGSDVDTSDESGTIEKADPTEGGGLLVIVYEAQDVEGKHTRQVYRIVRGER
KTKIKKNDPRMEQEFQVCEBPPINDMGOILEVISRPSIGHSNENIGYVISLAD
VNNKRINKYHLIDSKNRIOLELQWRS"
complement(join(28558..28797,29836..29946))
/gene="OJ1538_H05.8"
complement(join(28558..28797,29836..29946))
/gene="OJ1538_H05.8"
/note="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
complement(join(30919..31115,32919..32985,33026..33083,
34161..34364,34434..34933))
/gene="OJ1538_H05.9"
complement(join(30919..31115,32919..32985,33026..33083,
34161..34364,34434..34933))
/gene="OJ1538_H05.9"
```

```

gene      /note="polypeptidein-like"
          /join(37171..37359,37450..37458)
          /gene="OJ1538_H05.10"
          /join(37171..37359,37450..37458)
misc_feature
          /gene="OJ1538_H05.10"
          /note="hypothetical ORF
          predicted by GlimmerM
          this category is not included in IRGSP standard"
          complement(join(41460..41475,42412..42776))
gene      /gene="OJ1538_H05.11"
          /complement(join(41460..41475,42412..42776))
misc_feature
          /gene="OJ1538_H05.11"
          /note="hypothetical ORF
          predicted by GENSCAN
          this category is not included in IRGSP standard"
          44776..45989
          /gene="OJ1538_H05.12"
          /join(<44776..44829,45189..45989)
          /gene="OJ1538_H05.12"
          /note="start and end point are not identified"
          /join(44776..44829,45189..45989)

Query Match      83.2%; Score 20.8; DB 15; Length 132282;
Best Local Similarity 91.7%; Pred. No. 4.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAGCTTCATCT 24
DB      31570 AAAAAAAAACTAGCTGTAT 31547

RESULT 19
LOCUS      AC104103      137278 bp      DNA      linear      ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-37816 from chromosome 17, complete
sequence.
ACCESSION      AC104103
VERSION      AC104103.5 GI:33457243
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 137278)
Nguyen,C. and Bielicki,L.
The sequence of Mus musculus BAC clone RP24-37816
Unpublished (2001)
2 (bases 1 to 137278)
Wilson,R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 137278)
McPherson,J.D. and Waterston,R.H.
Direct Submision
Submitted (03-DEC-2001) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 137278)
Wilson,R.K.
Direct Submision
Submitted (15-MAY-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 137278)
Wilson,R.K.
Direct Submision
Submitted (06-AUG-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 137278)
Wilson,R.
Direct Submision
Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 6, 2003 this sequence version replaced gi:30725575.
COMMENT

```

```

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BB0378106
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RP24-37816 BAC library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC122502.

FEATURES
source
1..137278
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/map="17"
/clone="RP24-37816"
/clone_1b="RP24-37816"
235..318
/rpt_family="L1"
repeat_region
592..1235
/rpt_family="L1"
repeat_region
1383..1828
/rpt_family="L1"
repeat_region
4312..4662
/rpt_family="MALR"
repeat_region
4689..4841
/rpt_family="MIR"
repeat_region
5308..5597
/rpt_family="B4"
repeat_region
6999..7142
/rpt_family="B4"
repeat_region
7230..7945
/rpt_family="L1"
repeat_region
9588..9705
/rpt_family="B4"
repeat_region
10600..10830
/rpt_family="MALR"
repeat_region
10944..10998
/rpt_family="L2"
repeat_region
11030..11079
/rpt_family="ERV1"
repeat_region
11049..11095
/rpt_family="ERV1"
repeat_region
11286..11332

```

```

repeat_region      35584..37021      /rpt_family="L1"
repeat_region      38187..38389      /rpt_family="MER1_type"
repeat_region      38615..38847      /rpt_family="MALR"
repeat_region      39437..39751      /rpt_family="L1"
repeat_region      40806..40995      /rpt_family="B4"
repeat_region      41217..41349      /rpt_family="B4"
repeat_region      41866..41909      /rpt_family="MIR"

Query Match      83.2% ; Score 20.8 ; DB 9 ; Length 137278 ;
Best Local Similarity 91.7% ; Pred. No. 4.9e+02 ;
Matches 22 ; Conservative 0 ; Mismatches 2 ; Indels 0 ; Gaps 0 ;

Cy 1 AAAAAAAAACTATAGCTTGATCT 24
      |||||
Db 84745 AAAAAAAAACTAAGCTGATCT 84768

RESULT 20
AL533784/c
LOCUS
DEFINITION
169780 bp DNA linear PRI 18-MAY-2005
Human DNA sequence from clone RP11-314p6 on chromosome 10 contains
part of the PCDH15 gene for protocadherin 15 and the 5' end of a
novel gene, complete sequence.
AL533784
AL533784.15 GI:15142002
HTG: PCDH15.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 169780)
Tracey A.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Aug 9, 2001 this sequence version replaced gi:14455861.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-314p6 is from the library RPC11-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.hcm
VECTOR: pBAC3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30) ; an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.

Location/Qualifiers
1..169780

```

```

repeat_region      35584..37021      /rpt_family="L1"
repeat_region      38187..38389      /rpt_family="MER1_type"
repeat_region      38615..38847      /rpt_family="MALR"
repeat_region      39437..39751      /rpt_family="L1"
repeat_region      40806..40995      /rpt_family="B4"
repeat_region      41217..41349      /rpt_family="B4"
repeat_region      41866..41909      /rpt_family="MIR"

Query Match      83.2% ; Score 20.8 ; DB 9 ; Length 137278 ;
Best Local Similarity 91.7% ; Pred. No. 4,9e+02 ;
Matches 22 ; Conservative 0 ; Mismatches 2 ; Indels 0 ; Gaps 0 ;

Cy 1 AAAAAAAAACTATAGCTTGATCT 24
      |||||
Db 84745 AAAAAAAAACTAAGCTGATCT 84768

RESULT 20
AL533784/c
LOCUS
DEFINITION
169780 bp DNA linear PRI 18-MAY-2005
Human DNA sequence from clone RP11-314P6 on chromosome 10 contains
part of the PCDH15 gene for protocadherin 15 and the 5' end of a
novel gene, complete sequence.
AL533784
AL533784.15 GI:15142002
HTG: PCDH15.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 169780)
Tracey A.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Aug 9, 2001 this sequence version replaced gi:14455861.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-314P6 is from the library RPC11-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.hcm
VECTOR: pBAC3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30) ; an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.

Location/Qualifiers
1..169780

```

```
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-314P6"
/clone_1fb="RP11-11.2"

1
/misc_feature
/note="Clone left end: RP11-314P6"
join(69121..69207,151456..151566,AL391356.22:20803..22096,
AL391356.22:34245..34435,AL356114.12:6646..6762,
AL356114.12:7286..7785)
/misc_feature
/note="RP11-257114.1-001"
join(69121..69207,151456..151566,AL391356.22:20803..22096,
AL391356.22:34245..34435,AL356114.12:6646..6762,
AL356114.12:7286..7785)
/misc_feature
/note="RP11-257114.1-001"
/product="novel transcript"
/note="match: ESTs: Em:BG773153.1
match: CDNA: Em:BC040305.1"
88417
/misc_feature
/locus_tag="RP11-257114.1-001"
/note="tag=right end: RP11-533N7"
join(complement(AC051618.7:29803..30169),
complement(AL356114.12:15906..16024),
complement(110703..110768),
complement(AL360214.19:11300..11460),
complement(AC013737.5:202897..203052),
complement(AC013737.5:180142..180261),
complement(AC013737.5:163373..163483),
complement(AC013737.5:151048..151218),
complement(AC013737.5:70600..70708),
complement(AC024073.7:158363..158475),
complement(AC024073.7:140110..140316),
complement(AC024073.7:129561..129620),
complement(AC024073.7:127871..128020),
complement(AC024073.7:97527..97720),
complement(AC024073.7:77302..77434),
complement(AC024073.7:34411..34490),
complement(AC024073.7:23758..23851),
complement(AC024073.7:11184..11312),
complement(AC02671.9:67357..67662),
complement(AC02671.9:64657..64881),
complement(AC02671.9:40114..40230),
complement(AC016817.6:150055..150195),
complement(AC016817.6:148035..148147),
complement(AC016817.6:129169..129278),
complement(AC016817.6:127118..127258),
complement(AC016817.6:91546..91673),
complement(AC016817.6:54945..55160),
complement(AC016817.6:45478..45566),
complement(AC016817.6:28623..28799),
complement(AC016817.6:19618..19836),
complement(AC016817.6:16868..16876),
complement(AC016817.6:15696..15851),
complement(AL365496.15:167834..169916))
/gene="PCDH15"
/locus_tag="RP11-449J3.2-001"
join(complement(AC051618.7:29803..30066),
complement(AL356114.12:15906..16024),
complement(110703..110768),
complement(AL360214.19:11300..11460),
complement(AC013737.5:202897..203052),
complement(AC013737.5:180142..180261),
complement(AC013737.5:163373..163483),
complement(AC013737.5:151048..151218),
complement(AC013737.5:70600..70708),
complement(AC024073.7:158363..158475),
complement(AC024073.7:140110..140316),
complement(AC024073.7:129561..129695),
complement(AC024073.7:127871..128020),
complement(AC024073.7:97527..97720),
complement(AC024073.7:77302..77434),
complement(AC024073.7:34411..34490),
complement(AC024073.7:23758..23851),
complement(AC024073.7:11184..11312),
complement(AC02671.9:67357..67662),
complement(AC02671.9:64657..64881),
complement(AC02671.9:40114..40230),
complement(AC016817.6:150055..150195),
complement(AC016817.6:148035..148147),
complement(AC016817.6:129169..129278),
complement(AC016817.6:127118..127258),
complement(AC016817.6:91546..91673),
complement(AC016817.6:54945..45566),
complement(AC016817.6:28623..28799),
complement(AC016817.6:19618..19836),
complement(AC016817.6:16868..16876),
complement(AC016817.6:15696..15851),
complement(AL365496.15:167834..169916))
/gene="PCDH15"
/locus_tag="RP11-449J3.2-001"
/product="protocadherin 15"
/note="match: ESTs: Em:BG004571.1 Em:CF743213.1
match: CDNA: Em:AV029205.1 Em:AV029237.1"
join(complement(AL356114.12:15906..15996),
complement(110703..110768),
complement(AL360214.19:11300..11460),
```

mRNA

mRNA

CDS

```
complement(AC024073.7:23758..23851),
complement(AC024073.7:11184..11312),
complement(AC02671.9:67357..67662),
complement(AC02671.9:64657..64881),
complement(AC02671.9:39044..40230)
/gene="PCDH15"
/locus_tag="RP11-449J3.2-002"
join(complement(AC051618.7:29803..30066),
complement(AL356114.12:15906..16024),
complement(110703..110768),
complement(AL360214.19:11300..11460),
complement(AC013737.5:202897..203052),
complement(AC013737.5:180142..180261),
complement(AC013737.5:163373..163483),
complement(AC013737.5:151048..151218),
complement(AC013737.5:70600..70708),
complement(AC024073.7:158363..158475),
complement(AC024073.7:140110..140316),
complement(AC024073.7:129561..129695),
complement(AC024073.7:127871..128020),
complement(AC024073.7:97527..97720),
complement(AC024073.7:77302..77434),
complement(AC024073.7:34411..34490),
complement(AC024073.7:23758..23851),
complement(AC024073.7:11184..11312),
complement(AC02671.9:67357..67662),
complement(AC02671.9:64657..64881),
complement(AC02671.9:39044..40230)
/gene="PCDH15"
/locus_tag="RP11-449J3.2-001"
/product="protocadherin 15"
/note="match: CDNA: Em:AL83134.1"
join(complement(AC051618.7:29803..30169),
complement(AL356114.12:15906..16024),
complement(110703..110768),
complement(AL360214.19:11300..11460),
complement(AC013737.5:202897..203052),
complement(AC013737.5:180142..180261),
complement(AC013737.5:163373..163483),
complement(AC013737.5:151048..151218),
complement(AC013737.5:70600..70708),
complement(AC024073.7:158363..158475),
complement(AC024073.7:140110..140316),
complement(AC024073.7:129561..129695),
complement(AC024073.7:127871..128020),
complement(AC024073.7:97527..97720),
complement(AC024073.7:77302..77434),
complement(AC024073.7:34411..34490),
complement(AC024073.7:23758..23851),
complement(AC024073.7:11184..11312),
complement(AC02671.9:67357..67662),
complement(AC02671.9:64657..64881),
complement(AC02671.9:40114..40230),
complement(AC016817.6:150055..150195),
complement(AC016817.6:148035..148147),
complement(AC016817.6:129169..129278),
complement(AC016817.6:127118..127258),
complement(AC016817.6:91546..91673),
complement(AC016817.6:54945..45566),
complement(AC016817.6:28623..28799),
complement(AC016817.6:19618..19836),
complement(AC016817.6:16868..16876),
complement(AC016817.6:15696..15851),
complement(AL365496.15:167834..169916))
/gene="PCDH15"
/locus_tag="RP11-449J3.2-001"
/product="protocadherin 15"
/note="match: ESTs: Em:BG004571.1 Em:CF743213.1
match: CDNA: Em:AV029205.1 Em:AV029237.1"
join(complement(AL356114.12:15906..15996),
complement(110703..110768),
complement(AL360214.19:11300..11460),
```

```

complement (AC013737.5:202897..203052),
complement (AC013737.5:180142..180261),
complement (AC013737.5:163373..163483),
complement (AC013737.5:151048..151218),
complement (AC013737.5:70600..70708),
complement (AC024073.7:158353..158475),
complement (AC024073.7:140110..140316),

Query Match      83.2% Score 20.8; DB 8; Length 169780;
Best Local Similarity 91.7%; Pred. No. 4.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATGCTGATCT 24
    |||||
Db 50502 AAAAAAAAACTATGCTGATCT 50479

RESULT 21
BS000168      184554 bp      DNA      linear      PRI 12-JUN-2004
LOCUS      Pan troglodytes chromosome 22 clone:RP43-021P14, map 22, complete
DEFINITION
ACCESSION      BS000168 BA000046
VERSION      BS000168.2 GI:38016010
KEYWORDS      HTG.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
      Homidae; Pan.
REFERENCE
AUTHORS      1
TITLE      The International Chimpanzee Chromosome 22 Consortium.
JOURNAL      DNA sequence and comparative analysis of chimpanzee chromosome 22
REFERENCE      Nature 429, 382-388 (2004)
AUTHORS      2 (bases 1 to 184554)
      Wang, S., Cai, Z., Wang, B., Zheng, H., Zhang, Y., Zhang, X., Zhu, G.,
      Lu, G., Fu, G. and Chen, Z.
TITLE      Direct Submission.
JOURNAL      Submitted (26-MAY-2003) Shengyue Wang, Chinese National Human
      Genome Center at Shanghai, Genomic Sequencing; No.250 Bibo Road,
      Zhang Jiang Hi-TECH Park, Shanghai 201203, CHINA
      (E-mail:wangsy@chgc.sh.cn, URL:http://www.chgc.sh.cn,
      Tel:86-21-50801919, Fax:86-21-50801922)
      On Oct 28, 2003 this sequence version replaced gi:37537435.
COMMENT
      The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
      *Chinese National Human Genome Center at Shanghai, Shanghai, China;
      *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
      of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
      Center, Daejeon, Korea;
      *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
      *National Institute of Genetics, Mishima, Japan;
      *National Yang Ming University Genome Research Center, Taipei,
      Taiwan;
      *RIKEN Genomic Sciences Center, Yokohama, Japan.
      ----- Genome Center
      Center: Chinese National Human Genome Center at Shanghai Center
      code: CHGCs
      Web site: http://chgc.sh.cn
      Contact: wangsy@chgc.sh.cn
      ----- Project Information
      Center project name: The Chimpanzee Chromosome 22 Sequencing Project
      Center clone name: RP43-021P14
      ----- Summary Statistics
      Sequencing vector: pUC18, 100% of reads
      Chemistry: Dye-terminator Big Dye and ET, 100% of reads Assembly
      Program: Phrap; version 0.990329
      Consensus quality: 183135 bases at least Q40
      Consensus quality: 184057 bases at least Q20
      Consensus quality: 184465 bases at least Q20
      Quality coverage: 11.5x
      -----
      This sequence was finished as follows unless otherwise noted: all
      regions were double stranded, sequenced with an alternate

```

```

chemistry, or covered by high quality data (i.e., phred quality >=
30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.
-----
Source Information:
The RP43 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pBACe3.6
Sequence Quality Assessment:
This entry has been annotated with sequence
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in
10,000 bp.
-----
Neighboring clones: PTB-022L04 (left) and RP43-042E01 (right).
-----
FEATURES
source
location/Qualifiers
1..184554
/organism="Pan troglodytes".
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="RP43-021P14"
/clone_11b="RP43 chimpanzee BAC"

ORIGIN
Query Match      83.2% Score 20.8; DB 8; Length 184554;
Best Local Similarity 91.7%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATGCTGATCT 24
    |||||
Db 101170 AAAAAAAAACTATGCTGATCT 101193

RESULT 22
AC019179      196721 bp      DNA      linear      PRI 09-MAY-2001
LOCUS      Homo sapiens BAC clone RP11-240A16 from 4, complete sequence.
DEFINITION
ACCESSION      AC019179
VERSION      AC019179.4 GI:11120947
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
      Homidae; Homo.
REFERENCE
AUTHORS      1 (bases 1 to 196721)
      Sulston, J.E. and Waterston, R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE
AUTHORS      2 (bases 1 to 196721)
      Harkins, R., Maupin, R., Gregory, S., Coblitz, B. and Fleming, A.
TITLE      The sequence of Homo sapiens BAC clone RP11-240A16
JOURNAL      Unpublished
REFERENCE
AUTHORS      3 (bases 1 to 196721)
      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (30-DEC-1999) Genome Sequencing Center, Washington
      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
      MO 63108, USA
      4 (bases 1 to 196721)
      Waterston, R.
REFERENCE
AUTHORS      Submitted (08-NOV-2000) Department of Genetics, Washington
      JOURNAL

```

REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 196721)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 8, 2000 this sequence version replaced gi:7630907.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0240A16

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, B.,
Tateiro, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP11-240A16;
actual end is at base position 196721 of RP11-240A16.

The sequence H_NH0240A16 from base position 157677 to 158503
contains a tandem repeat. The assembly is consistent with digest
information but the sequence fidelity cannot be guaranteed.
Location/Qualifiers
1. 196721
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-240A16"
/clone_1lb="RPCI-11"
3448..3840
/rpt_family="ERV"
4103..4596
/rpt_family="ERVK"
5164..5231
/rpt_family="MIR"
5616..5734
/rpt_family="MIR"
5861..5911
/rpt_family="Mariner"
6619..6749
/rpt_family="MaLR"
repeat_region

repeat_region 6727..6776
/rpt_family="MaLR"
repeat_region 8458..8684
/rpt_family="Alu"
repeat_region 8685..8843
/rpt_family="Alu"
repeat_region 10345..10912
/rpt_family="MaLR"
repeat_region 10952..11592
/rpt_family="L2"
repeat_region 14824..15153
/rpt_family="L1"
repeat_region 16195..16619
/rpt_family="ERVL"
repeat_region 16649..16822
/rpt_family="MaLR"
repeat_region 17787..17850
/rpt_family="Alu"
repeat_region 18581..18773
/rpt_family="MER1_type"
repeat_region 19637..19697
/rpt_family="L2"
repeat_region 19798..19971
/rpt_family="Alu"
repeat_region 19991..20128
/rpt_family="Alu"
repeat_region 22389..22580
/rpt_family="MIR"
repeat_region 23974..24287
/rpt_family="Alu"
repeat_region 25165..25399
/rpt_family="MIR"
repeat_region 25434..25719
/rpt_family="Alu"
repeat_region 26014..26157
/rpt_family="L2"
repeat_region 28448..28553
/rpt_family="MIR"
repeat_region 29093..29243
/rpt_family="Alu"
repeat_region 29715..30240
/rpt_family="MaLR"
repeat_region 30717..30823
/rpt_family="L1"
repeat_region 30859..30929
/rpt_family="L1"
repeat_region 30935..31516
/rpt_family="ERV1"
repeat_region 31502..31764
/rpt_family="ERV1"
repeat_region 31871..31981
/rpt_family="L1"
repeat_region 31982..32494
/rpt_family="ERVK"
repeat_region 32495..32523
/rpt_family="L1"
repeat_region 32579..32646
/rpt_family="MaLR"
repeat_region 32775..32922
/rpt_family="ERVL"
repeat_region 33108..33153
/rpt_family="ERVL"
repeat_region 33187..33486
/rpt_family="Alu"
repeat_region 33588..33696
/rpt_family="MaLR"
repeat_region 34154..34461
/rpt_family="L2"
repeat_region 34462..35265
/rpt_family="ERV1"
repeat_region 35266..35432
/rpt_family="L2"
repeat_region 35517..35685

```

repeat_region      /rpt_family="MALR"
35868      .35939
repeat_region      /rpt_family="MIR"
38314      .38622
repeat_region      /rpt_family="L1"
38628      .38765
repeat_region      /rpt_family="L1"
38872      .39419
repeat_region      /rpt_family="L1"
41647      .41946
repeat_region      /rpt_family="ALU"
42051      .42251
repeat_region      /rpt_family="L1"
42265      .42886
repeat_region      /rpt_family="ERV1"
42887      .43631
repeat_region      /rpt_family="L1"
43803      .44108
repeat_region      /rpt_family="ALU"
47859      .47918
repeat_region      /rpt_family="ERV1"
47963      .48092
repeat_region      /rpt_family="ERV1"
48723      .49085
repeat_region      /rpt_family="ERV1"

Query Match      83.2% Score 20.8; DB 8; Length 196721;
Best Local Similarity 91.7%; Pred. No. 4.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAGCTGATCT 24
Db      91 AAAAAAAAAAGCTAAAGCTGATCT 114

RESULT 23
AC105245      200721 bp      DNA      linear      PRI 15-MAR-2002
LOCUS      Homo sapiens chromosome 18, clone CTD-2515C13, complete sequence.
AC105245
VERSION      AC105245.4 GI:19482347
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 200721)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone CTD-2515C13
2 (bases 1 to 200721)
Unpublished
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguski,V., Bouckgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,Y., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,
Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,T., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

```

```

TITLE      Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL      Direct Submission
REFERENCE      Submitted (26-DEC-2001) Whitehead Institute/MIT Center for Genome
AUTHORS      Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 200721)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguski,V., L.,
Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Labocque,K., Lamarez,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,T., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 15, 2002 this sequence version replaced gi:18598792.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23209
Center clone name: 2515_C13
-----
Location/Qualifiers
1. 200721
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone_1lb="CTD2 Human BAC"
746. 835
/rpt_family="MIR"
1093. 1253
/rpt_family="MIR"
1256. 1283
/rpt_family="AT-rich"
Complement(1295..1507)
/rpt_family="MIR"
Complement(4122..4750)
/rpt_family="LIME3"
Complement(4772..4957)
/rpt_family="LIME3"
Complement(5672..6016)
/rpt_family="LIPB3"
6017. 6050
/rpt_family="(TG)n"
Complement(6051..6705)
/rpt_family="LIPB3"
Complement(6710..8029)
/rpt_family="LIM1"
9321. 9397

```

```
repeat_region /rpt_family="AT_rich"
10279. .10338
repeat_region /rpt_family="(TG)n"
complement(10339. .10696)
repeat_region /rpt_family="MLT1A1"
11022. .11144
repeat_region /rpt_family="MIR"
11472. .11717
repeat_region /rpt_family="MIR"
12228. .12257
repeat_region /rpt_family="(TG)n"
12338. .12618
repeat_region /rpt_family="AluSc"
12623. .12714
repeat_region /rpt_family="AluY"
12837. .13151
repeat_region /rpt_family="MLT1A1"
13715. .14154
repeat_region /rpt_family="L2"
14159. .14282
repeat_region /rpt_family="L2"
14159. .14282
repeat_region /rpt_family="(5752. .15744)
complement(15752. .16142)
repeat_region /rpt_family="MER66A"
16267. .16393
repeat_region /rpt_family="(CATATA)n"
16513. .16821
repeat_region /rpt_family="AluY"
17343. .17364
repeat_region /rpt_family="AT_rich"
complement(17612. .17828)
repeat_region /rpt_family="LIME"
17904. .17973
repeat_region /rpt_family="T-rich"
18175. .18215
repeat_region /rpt_family="AT_rich"
18398. .18437
repeat_region /rpt_family="(CTATT)n"
18440. .18549
repeat_region /rpt_family="(CCTAT)n"
18749. .18770
repeat_region /rpt_family="AT_rich"
complement(19654. .19759)
repeat_region /rpt_family="AluJo"
19760. .19803
repeat_region /rpt_family="(TAA)n"
complement(19804. .19919)
repeat_region /rpt_family="AluJo"
20158. .20182
repeat_region /rpt_family="AT_rich"
20329. .20377
repeat_region /rpt_family="AT_rich"
20383. .20880
repeat_region /rpt_family="MLT2B2"
complement(21624. .21765)
repeat_region /rpt_family="LIMEC"
22916. .22943
repeat_region /rpt_family="AT_rich"
23785. .23870
repeat_region /rpt_family="(TA)n"
complement(24156. .24703)
repeat_region /rpt_family="MLT1H1"
complement(25380. .25464)
repeat_region /rpt_family="MIR3"
complement(25571. .26034)
repeat_region /rpt_family="MLT1B"
complement(26035. .26395)
repeat_region /rpt_family="TBRIC"
complement(26396. .26465)
repeat_region /rpt_family="MLT1B"
26486. .26521
repeat_region /rpt_family="AT_rich"
```

```
repeat_region 27134. .27171
repeat_region /rpt_family="(CATATA)n"
28254. .28370
repeat_region /rpt_family="MLT1B"
28634. .28953
repeat_region /rpt_family="AluSc"
complement(29285. .29436)
repeat_region /rpt_family="AluJo"
30753. .30823
repeat_region /rpt_family="Tiger3 (Golem)"
30825. .30958
repeat_region /rpt_family="Tiger3 (Golem)"
complement(31053. .31419)
repeat_region /rpt_family="MER82"
31470. .31551
repeat_region /rpt_family="Tiger5"
complement(31552. .31827)
repeat_region /rpt_family="MER82"

Query Match 83.2% Score 20.8; DB 8; Length 200721;
Best Local Similarity 91.7%; Pred. No. 4.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATGCTGATCT 24
DB 150390 AGAAAAAAATATAGCTTGATCT 150413

RESULT 24
AC121626/c 220407 bp DNA linear HTG 15-NOV-2002
LOCUS AC121626
DEFINITION Rattus norvegicus clone CH230-92A9, WORKING DRAFT SEQUENCE.
ACCESSION AC121626
VERSION AC121626.4 GI:25008835
KEYWORDS HTG: HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 220407)
REFERENCE
1 Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Dlyva, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebreyes, E., Geer, K., Gill, R., Garcia, A., Garrow, M., Gervais, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hu, Y. S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyan, C.,
Kows, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, K., Martinez, E.,
Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, S.,
Milesavjevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokeneme, O., Okunnu, G., Olarinmoyegun, A., Pal, S., Parks, K.,
```


Pastermak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
 Pucato, M., Quito, J., Rachlin, E., Reeves, K., Reister, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
 Stealm, E., Strong, R., Sutton, A., Svatek, A., Taboi, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tinney, A., Trejos, Z., Uman, K.,
 Valdes, R., Verz, V., Villanana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederstockern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weiss, R.A. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 220407)
 Worley, K.C.
 Direct Submission
 Submitted (21-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 220407)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:23194733.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWSN
 Center clone name: CH230-92A9
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 210154 bases at least Q40
 Consensus quality: 211654 bases at least Q30
 Consensus quality: 213068 bases at least Q20
 Estimated insert size: 217760; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 220407; contig of 220407 bp in length.
 Location/Qualifiers

```

source      1. .220407
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-92A9"
            1. .1524
misc_feature clone_end=Sp6
             end_sequence:BH292498"
misc_feature clone_end:Sp6"
             /note="wgs_end_extension
             3617. .3990
             /note="clone boundary
             clone_end:Sp6
             site:
             end_sequence:BH292497"
misc_feature 163773. .165622
             /note="wgs_contig"
misc_feature 167304. .167741
             /note="clone boundary
             clone_end:T7
             site:
             end_sequence:BH292497"
misc_feature 167325. .167786
             /note="clone boundary
             clone_end:T7
             site:
             end_sequence:BH292497"

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 220407;
Best Local Similarity 91.7%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      1 AAAAAAAAACTATAGCTGTGATCT 24
        ||||||| |||||||||
Db      37600 AAAAAAAAAAAATGCTTGATCT 37577

RESULT 25
AC099385/c AC099385      222358 bp   DNA       linear    HTG 13-MAY-2003
LOCUS      Rattus norvegicus clone CH230-182G12, WORKING DRAFT SEQUENCE.
DEFINITION AC099385
ACCESSION  AC099385.5 GI:30580043
VERSION     HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS    Rattus norvegicus (Norway rat)
SOURCE      Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
            1 (bases 1 to 222358)
AUTHORS     Muzny,D.,Marler,M.,Metzker,M.,Lee,A.,Abramson,S.,Adams,C.,Alder,J.,
            Allen,C.,Allen,H.,Alshbrook,S.,Amin,A.,Angulano,D.,
            Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,
            Baldwin,D.,Bandaranaikhe,D.,Barber,M.,Barnstead,M.,Benahmed,F.,
            Biswal,K.,Blair,J.,Blankenburg,K.,Blych,P.,Brown,M.,
            Bryant,N.,Buhaq,C.,Burck,P.,Butrell,K.,Calderon,E.,
            Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,R.,Center,A.,
            Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
            Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,
            Davila,M.L.,Davis,C.,Davy-Carroil,L.,De Anda,C.,Dederich,D.,
            Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K.,
            Draper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duvál,B.,Eaves,K.,
            Bean,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,
            Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P.,
            Fraser,C.M.,Gabisi,A.,Ganta,R.,García,A.,Garner,T.,Garza,M.,
            Gebregeorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guerrero,A.,
            Guinarte,P.,Haaland,W.,Hamill,C.,Hamilton,C.,Hamilton,K.,
            Harvey,Y.,Haylak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
            Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,
            Hollins,B.,Howells,S.,Huylx,S.,Hume,J.,Idelblid,T.,Jackson,A.,
            Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jollivet,A.,
            Karpathy,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Kovar,C.,
            Kowis,C.,Kraft,C.L.,Ledoux,H.,Leyvan,J.,Lewis,L.,Li,Z.,Liu,J.,
            Liu,J.,Liu,M.,Liu,Y.,London,P.,Longacre,S.,Lopez,J.,

```

Lorenshewu, L., Louisege, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, B.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwackemeke, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,
 Valae, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Woodin, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, D., von
 Niederhuesen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weierhach, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 222358)
 Worley, K.C.
 Direct Submission
 Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 222358)
 Rat Genome Sequencing Consortium.
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 13, 2003 this sequence version replaced g1:23269089.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Project name: GKDB
 Center project name: GKDB
 Center clone name: CH230-182G12
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 207245 bases at least Q40
 Consensus quality: 209217 bases at least Q30
 Consensus quality: 210807 bases at least Q20
 Estimated insert size: 216298; sum-of-contigs estimation
 Quality coverage: 10x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the size
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 1 222358: contig of 222358 bp in length.
 Location/Qualifiers
 source
 1..222358
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-182G12"
 misc_feature
 1..2454
 /note="wgs contig"
 misc_feature
 7336..8612
 /note="wgs contig"
 misc_feature
 /note="wgs contig"

ORIGIN
 Query Match 83.2% Score 20.8; DB 14; Length 222358;
 Best Local Similarity 91.7%; Pred. No. 4,2e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 AAAAAAAAACTAGCTGATCTT 25
 Db 120902 AAAAAAAAAAGATGCTTATCTT 120879

RESULT 26
 AC125295/c
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-9A11, WORKING DRAFT SEQUENCE, 5
 unoriented pieces.
 ACCESSION
 AC125295
 VERSION
 AC125295.3 GI:24817960
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS
 Rattus norvegicus (Norway rat)
 SOURCE
 Rattus norvegicus
 ORGANISM
 Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.
 1 (bases 1 to 222669)
 1
 Muzny, D., Marie, Metzger, M., Lee, Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsdorfs, S., Amin, A., Anguiano, D.,
 Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F.,
 Bismah, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Buttrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Greengrads, E., Geer, K., Gill, R., Girdley, M., Guerra, M., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, J., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshewu, L., Louisege, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, B.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 54762: contig of 54762 bp in length
 * 54763 54862: gap of unknown length

FEATURES	source	location/Qualifiers
*	54853	146416: contig of 91554 bp in length
*	146417	146516: gap of unknown length
*	146517	220312: contig of 73796 bp in length
*	220313	220410: gap of unknown length
*	220413	221430: contig of 1018 bp in length
*	221431	221530: gap of unknown length
*	221531	222669: contig of 1139 bp in length.
FEATURES		
source		1. 222669
		/organism="Rattus norvegicus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10116"
		/clone="CH230-9a11"
		/complement(8912..9573)
		/note="clone boundary
		clone end:Sp6
		site:ECORI
		end_sequence:BH338890"
		51684..52802
		/note="wgs_end_extension
		clone_end:Sp6"
		53655..54762
		/note="wgs_end_extension
		clone_end:Sp6"
		54763..54862
		/estimated_length=unknown
		146417..146516
		/estimated_length=unknown
		146517..147699
		/note="wgs_end_extension
		clone_end:Sp6"
		149367..151085
		/note="wgs_end_extension
		clone_end:Sp6"
		15136..152459
		/note="wgs_end_extension
		clone_end:Sp6"
		152510..153349
		/note="wgs_end_extension
		clone_end:Sp6"
		220313..220412
		/estimated_length=unknown
		221431..221530
		/estimated_length=unknown
ORIGIN		
Query Match	83.2%; Score 20.8; DB 14; Length 222669;	
Best Local Similarity	91.7%; Pred. No. 4.2e+02;	
Matches	22; Conservative 0; Pident 2; Indels 0; Gaps 0;	
QY	1 AAAAAAAAAACTATAGCTTGATCT 24	
Db	19017 AAAAAAAAAAATAGCTTGATCT 18994	
RESULT 27		
LOCUS	AC123870	223132 bp DNA linear ROD 08-NOV-2003
DEFINITION	Mus musculus BAC clone RP23-246A19 from 15, complete sequence.	
ACCESSION	AC123870	
VERSION	AC123870.3	GI:29164698
KEYWORDS	HTG.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 223132)	

AUTHORS Nguyen,C., Creason,K. and Bielicki,L.
TITLE The sequence of Mus musculus BAC clone RP23-246A19
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 223132)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 223132)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (01-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 223132)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 223132)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (23-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 223132)
AUTHORS Wilson,R.
TITLE Direct Submision
JOURNAL Submitted (08-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 23, 2003 this sequence version replaced gi:22475512.
----- Genome Center
----- Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
Center project name: M_BA0246A19

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
and Minko Tateno in the laboratory of Pieter de Jong
(http://www.chori.org) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (http://www.resgen.com) or
Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1..223132
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"

/clone="RP23-246A19"
/clone_11b="RPCI-23"
repeat_region
/rpt_family="B2"
686..859
repeat_region
/rpt_family="A1u"
1084..1232
repeat_region
/rpt_family="A1u"
1368..1500
repeat_region
/rpt_family="A1u"
1630..1734
repeat_region
/rpt_family="MIR"
2418..2539
repeat_region
/rpt_family="MER1_type"
2587..2754
repeat_region
/rpt_family="B4"
2805..2975
repeat_region
/rpt_family="B2"
4629..4726
repeat_region
/rpt_family="L1"
4777..4875
repeat_region
/rpt_family="BCL_MM"
4965..5112
repeat_region
/rpt_family="A1u"
5232..5321
repeat_region
/rpt_family="B4"
5294..5346
repeat_region
/rpt_family="A1u"
5599..5672
repeat_region
/rpt_family="MER1_type"
5851..5982
repeat_region
/rpt_family="A1u"
6212..6359
repeat_region
/rpt_family="A1u"
6626..6751
repeat_region
/rpt_family="B4"
7352..7462
repeat_region
/rpt_family="A1u"
7651..7803
repeat_region
/rpt_family="A1u"
8055..8099
repeat_region
/rpt_family="B4"
8100..8175
repeat_region
/rpt_family="A1u"
11025..11182
repeat_region
/rpt_family="B4"
11061..11183
repeat_region
/rpt_family="A1u"
12340..12421
repeat_region
/rpt_family="B4"
12425..12442
repeat_region
/rpt_family="A1u"
15325..15459
repeat_region
/rpt_family="B4"
15804..15915
repeat_region
/rpt_family="A1u"
16085..16198
repeat_region
/rpt_family="B2"
16346..16451
repeat_region
/rpt_family="B2"
16830..16981
repeat_region
/rpt_family="B4"
17055..17190
repeat_region
/rpt_family="A1u"
17279..17382
repeat_region
/rpt_family="A1u"
17407..17592
repeat_region
/rpt_family="MER1_type"
17985..18130
repeat_region
/rpt_family="B4"
18131..18281
repeat_region
/rpt_family="A1u"
18282..18361
repeat_region
/rpt_family="B4"
18580..18662
repeat_region

```

repeat_region      /rpt_family="Alu"
19316..19619
/rpt_family="B4"
repeat_region      /rpt_family="B4"
19898..20030
/rpt_family="CR1"
repeat_region      /rpt_family="CR1"
20675..20910
/rpt_family="ERVK"
repeat_region      /rpt_family="CR1"
22708..22945
/rpt_family="CR1"
repeat_region      /rpt_family="MIR"
24063..24090
/rpt_family="MIR"
repeat_region      /rpt_family="ID"
24613..24689
/rpt_family="ID"
repeat_region      /rpt_family="Alu"
29919..30065
/rpt_family="Alu"
repeat_region      /rpt_family="Alu"
30163..30310
/rpt_family="Alu"
repeat_region      /rpt_family="B2"
30375..30576
/rpt_family="B2"
repeat_region      /rpt_family="B2"
30599..30787
/rpt_family="B2"
repeat_region      /rpt_family="B2"
30802..31028
/rpt_family="B2"
repeat_region      /rpt_family="B2"
31146..31260
/rpt_family="Alu"
repeat_region      /rpt_family="B2"
31315..31505
/rpt_family="B2"
repeat_region      /rpt_family="B2"
31331..31393
/product="CRNA-Ser"
/crna="CRNA-Ser"
/note="Likely pseudogene (HMM Sc=32.49 / Sec struct
Sc=11.21)"
repeat_region      /rpt_family="ID"
31557..31636
/rpt_family="ID"
crna               /product="CRNA-Val"
31563..31634
/note="Likely pseudogene (HMM Sc=20.22 / Sec struct
Sc=0.91)"
repeat_region      /rpt_family="B2"
31670..31869
/rpt_family="B2"
repeat_region      /rpt_family="Alu"
31888..32005
/rpt_family="Alu"
repeat_region      /rpt_family="B4"
32006..32031
Query Match      83.2% Score 20.8; DB 9; Length 223132;
Best Local Similarity 91.7%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAACTATGCTTGATCT 24
Db 136473 AAAAAAAAACTATGCTTGCTT 136450

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT

```

Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gburegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hernandez, J., Harvey, Y., Havlak, P., Hayes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hu, Y., Hume, J., Idubird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Lervan, J., Lewis, L., Li, Z., Liu, J., Liu, D., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, D., Lorenshewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedi, O., Okunolu, G., Olarnpunsoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Pioppert, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaiz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanli, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 228995)

Direct Submission
Submitted (29-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 228995)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (03-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 14, 2002 this sequence version replaced gi:2277017.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

-----Genome Center-----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
-----Project Information-----
Center project name: GPKY
Center clone name: CH230-968

```

----- Summary Statistics
Assembly program: Phrap version 0.990329
Consensus quality: 203410 bases at least Q40
Consensus quality: 205983 bases at least Q30
Consensus quality: 207169 bases at least Q20
Estimated insert size: 225230; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 227616: contig of 227616 bp in length
* 227617 227716: gap of unknown length
* 227717 228995: contig of 1279 bp in length.
Location/Qualifiers
1. 228995
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-968"
1. 1108
/feature="wgs_contig"
22618. 223794
/feature="wgs_contig"
223845. 225633
/feature="wgs_contig"
226135. 227616
/feature="wgs_contig"
227617. 227716
/feature="wgs_contig"
227617. 227716
/estimated_length=unknown

ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 228995;
Best Local Similarity 91.7%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTACTGATCT 24
38938 AAAAAAAAAAAATGCTGATCT 38961

RESULT 29
AC095612 248526 bp DNA linear HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-8B4, WORKING DRAFT SEQUENCE, 5
unordered pieces.
AC095612
AC095612.7 GI:30467184
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 248526)
Muzny, D., Mathe, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alstrooms, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Blewett, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Casati, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Crease, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
Delgado, O., Denison, S., Detamo, C., Ding, Y., Dinh, H., Diya, K.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
JOURNAL
AUTHORS
TITLE
JOURNAL
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24817800.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCVC
Center clone name: CH230-8B4
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P.,
Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgopoulos, E., Geer, K., Gill, R., Grady, M., Guertler, W., Guevara, W.,
Gunatiranga, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozano, R.J., Lu, X., Ma, U.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackeleme, O., Okunolu, G., Olarnunagsoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plommer, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.U.,
Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, M., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smjels, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tinney, A., Treloar, C.,
Vais, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Wolley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 248526)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 248526)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```

----- Summary Statistics

Assembly program: Atlas;
 Consensus quality: 238152 bases at least Q40
 Consensus quality: 239694 bases at least Q30
 Consensus quality: 240759 bases at least Q20
 Estimated insert size: 254402; sum-of-coverage estimation
 Quality coverage: 8x in Q40 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * The file is a "working draft" sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 239249: contig of 239249 bp in length
 * 239250 239349: gap of unknown length
 * 239350 241101: contig of 1752 bp in length
 * 241102 241201: gap of unknown length
 * 241202 243509: contig of 2308 bp in length
 * 243510 243609: gap of unknown length
 * 243610 245034: contig of 1225 bp in length
 * 245035 245134: gap of unknown length
 * 245135 248526: contig of 3392 bp in length.

FEATURES

source

1. 248526
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-8B4"

misc_feature

1. 1486

misc_feature

/note="wgs_end_extension
 clone_end:Sp6"
 complement(6321..7211)
 /note="clone_boundary
 clone_end:Sp6
 site:ECORI
 end_sequence:BH343826"
 complement(236139..236945)
 /note="clone_boundary
 clone_end:T7
 site:ECORI
 end_sequence:BH343824"

misc_feature

/note="wgs_end_extension
 clone_end:Sp6"
 complement(6321..7211)
 /note="clone_boundary
 clone_end:Sp6
 site:ECORI
 end_sequence:BH343826"
 complement(236139..236945)
 /note="clone_boundary
 clone_end:T7
 site:ECORI
 end_sequence:BH343824"

gap

239250..239349
 /estimated_length=unknown

gap

241102..241201
 /estimated_length=unknown

gap

243510..243609
 /estimated_length=unknown

gap

245035..245134
 /estimated_length=unknown

ORIGIN

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 14; Length 248526;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATGCTTGAATCTT 25
 DB 204615 AAAAAAAAAACGATGCTTCACTT 204638

RESULT 30
 AC103260/C

ACCESSION AC103260
 VERSION AC103260.6 GI:30578925

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.
 1 (bases 1 to 249383)

Muzny,D,Marle,M,McKer,M,Lee,A,Abrahamson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooke,S, Amin,A, Anguiano,D, Ayalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,K, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Bunay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,R, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flaggy,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garzara,M, Gbureggie,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W, Gunatane,P, Haaland,M, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Hulyc,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovat,C, Kowis,C, Kraft,C,L, Ledow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensunewa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapa,P, Martin,K, Martin,M, Martinez,E, Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minga,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Naoekelmech,O, Okunou,G, Olarnungsoot,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Plopper,F, Polindexter,A, Popovic,D, Primus,E, Pu,L, L, Piazzi,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajic,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Soes,J, Steinle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, Wiley,F, Williams,G, Willson,R, Wleczky,R, Wooden,H, Wortley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,U, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R,A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 13, 2003 this sequence version replaced gi:25007931.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center Project name: GPRS
Center clone name: CH230-48J15

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 210512 bases at least Q40
Consensus quality: 215471 bases at least Q30
Consensus quality: 215471 bases at least Q20
Estimated insert size: 237620; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Gendank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 33199: contig of 33199 bp in length
* 33200 33299: gap of unknown length
* 33300 252357: contig of 219058 bp in length
* 252358 252457: gap of unknown length
* 252458 253654: contig of 1197 bp in length
* 253655 253754: gap of unknown length
* 253755 254746: contig of 1492 bp in length.

source

1. 255246
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-48J15"

misc_feature

1. 2071
/note="wgs_end_extension
clone_end:Sp6"

misc_feature

2272. 3113
/note="clone boundary
clone_end:Sp6
site:EcoRI"

misc_feature

2573. 4090
end_sequence: BH274898"

misc_feature

/note="wgs contig"
10987. 11570
/note="clone boundary
clone_end:T7
site:EcoRI"

misc_feature

33200. 33299
end_sequence: BH274896"

gap

33300. 34958
/estimated_length=unknown

misc_feature

33300. 34958
/note="wgs_end_extension
clone_end:T7"

gap

252358. 252457
/estimated_length=unknown

gap

253655. 253754
/estimated_length=unknown

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 255246;
Best Local Similarity 91.7%; Pred. No. 4.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAACTAGCTGATCT 24
|||||

Db 85270 AAAAAAAAACTAGCTGATCT 85293

RESULT 32
AC103272

LOCUS
DEFINITION

AC103272 258727 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-230B2, WORKING DRAFT SEQUENCE, 7
unordered pieces.

ACCESSION

AC103272.8 GI:30578985

VERSION

HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Allen, C., Allen, H., Albrooke, S., Amin, A., Anguiano, D.,
Anyalebech, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kows, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muijase, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackelmehe, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C.,
Plopper, F., Polidexter, A., Popovic, D., Prims, E., Pu, L., L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Rulz, S. J.,
Sanders, M., Savery, G., Scherer, S., Scott, G., Shutsman, S., Shen, H.,
Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhou, X., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE
AUTHORS
TITLE
JOURNAL

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 258727)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 13, 2003 this sequence version replaced gi:23610443.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GURB
Center clone name: CH230-230B2

----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 252946 bases at least Q40
Consensus quality: 45518 bases at least Q30
Contig coverage: 264673; sum-of-contigs estimation
Estimated insert size: 264673; sum-of-contigs estimation
Quality coverage: 13x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 250595: contig of 250595 bp in length
* 250596 250695: gap of unknown length
* 250696 251720: contig of 1025 bp in length
* 251721 251820: gap of unknown length
* 251821 252830: contig of 1010 bp in length
* 252831 252930: gap of unknown length
* 252931 254058: contig of 1128 bp in length
* 254059 254158: gap of unknown length
* 254159 255419: contig of 1261 bp in length
* 255420 255519: gap of unknown length
* 255520 256839: contig of 1320 bp in length
* 256840 256939: gap of unknown length
* 256940 258727: contig of 1788 bp in length.

FEATURES

source

1. 258727
/organism="Rattus norvegicus"
/mol_type="Genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-230B2"

misc_feature

1. 2720
/note="wgs_end_extension"

misc_feature

4124..4469
/note="clone boundary"
clone_end:Sp6
site:EcoRI

misc_feature

end_sequence:B2111765"
complement(148864..250172)
/note="clone boundary"
clone_end:T7
site:EcoRI

gap

250596..250695
/estimated_length=unknown

gap

251721..251820
/estimated_length=unknown

gap

252831..252930
/estimated_length=unknown

gap

254059..254158
/estimated_length=unknown

gap

255420..255519
/estimated_length=unknown

gap

256840..256939
/estimated_length=unknown

ORIGIN

Query Match

Best Local Similarity 91.7%; Pred. No. 4.1e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy

2 AAAAAAAAACTTACCTGATCTT 25

Db

22273 AAAAAAAAAAGATGCTTCATCTT 22296

RESULT 33

CR847855

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

Project Information

Center project name: BMA385C13

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 94638 bases at least Q40

Consensus quality: 94657 bases at least Q30

Consensus quality: 94665 bases at least Q20

Insert size: 276749; sum-of-contigs

Insert size: 99880; 2.8% error; agarose-fp

Quality coverage: 4.39x in Q20 bases; sum-of-contigs Quality

coverage: 12.35x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```
* be preserved.
* 1 94669: contig of 94669 bp in length
* 94670 94769: gap of 100 bp
* 94770 98769: contig of 4000 bp in length
* 98770 98869: gap of 100 bp
* 98870 102869: contig of 4000 bp in length
* 102870 102969: gap of 100 bp
* 102970 106969: contig of 4000 bp in length
* 106970 107069: gap of 100 bp
* 107070 111069: contig of 4000 bp in length
* 111070 111169: gap of 100 bp
* 111170 115169: contig of 4000 bp in length
* 115170 115269: gap of 100 bp
* 115270 119269: contig of 4000 bp in length
* 119270 119369: gap of 100 bp
* 119370 123469: contig of 4000 bp in length
* 123470 127469: gap of 100 bp
* 127470 127569: contig of 4000 bp in length
* 127570 131569: gap of 100 bp
* 131570 131669: contig of 4000 bp in length
* 131670 135669: gap of 100 bp
* 135670 135769: contig of 4000 bp in length
* 135770 139769: gap of 100 bp
* 139770 139869: contig of 4000 bp in length
* 139870 143869: gap of 100 bp
* 143870 143969: contig of 4000 bp in length
* 143970 147969: gap of 100 bp
* 147970 148069: contig of 4000 bp in length
* 148070 152069: gap of 100 bp
* 152070 152169: contig of 4000 bp in length
* 152170 156169: gap of 100 bp
* 156170 156269: contig of 4000 bp in length
* 156270 160269: gap of 100 bp
* 160270 160369: contig of 4000 bp in length
* 160370 164369: gap of 100 bp
* 164370 164469: contig of 4000 bp in length
* 164470 168469: gap of 100 bp
* 168470 168569: contig of 4000 bp in length
* 168570 172569: gap of 100 bp
* 172570 172669: contig of 4000 bp in length
* 172670 176669: gap of 100 bp
* 176670 176769: contig of 4000 bp in length
* 176770 180769: gap of 100 bp
* 180770 180869: contig of 4000 bp in length
* 180870 184869: gap of 100 bp
* 184870 184969: contig of 4000 bp in length
* 184970 188969: gap of 100 bp
* 188970 189069: contig of 4000 bp in length
* 189070 193069: gap of 100 bp
* 193070 193169: contig of 4000 bp in length
* 193170 197169: gap of 100 bp
* 197170 197269: contig of 4000 bp in length
* 197270 201269: gap of 100 bp
* 201270 201369: contig of 4000 bp in length
* 201370 205369: gap of 100 bp
* 205370 205469: contig of 4000 bp in length
* 205470 209469: gap of 100 bp
* 209470 209569: contig of 4000 bp in length
* 209570 213569: gap of 100 bp
* 213570 213669: contig of 4000 bp in length
* 213670 217669: gap of 100 bp
* 217670 217769: contig of 4000 bp in length
* 217770 221769: gap of 100 bp
* 221770 221869: contig of 4000 bp in length
* 221870 225869: gap of 100 bp
* 225870 225969: contig of 4000 bp in length
* 225970 230069: gap of 100 bp
* 230070 234069: contig of 4000 bp in length
* 234070 234169: gap of 100 bp
* 234170 238169: contig of 4000 bp in length
* 238170 238269: gap of 100 bp
```

```
* 238270 242269: contig of 4000 bp in length
* 242270 242369: gap of 100 bp
* 242370 246369: contig of 4000 bp in length
* 246370 246469: gap of 100 bp
* 246470 250469: contig of 4000 bp in length
* 250470 250569: gap of 100 bp
* 250570 254569: contig of 4000 bp in length
* 254570 254669: gap of 100 bp
* 254670 258669: contig of 4000 bp in length
* 258670 258769: gap of 100 bp
* 258770 262769: contig of 4000 bp in length
* 262770 262869: gap of 100 bp
* 262870 266869: contig of 4000 bp in length
* 266870 266969: gap of 100 bp
* 266970 270969: contig of 4000 bp in length
* 270970 271069: gap of 100 bp
* 271070 275069: contig of 4000 bp in length
* 275070 275169: gap of 100 bp
* 275170 279169: contig of 4000 bp in length
* 279170 279269: gap of 100 bp
* 279270 281349: contig of 2080 bp in length.
* Location/Qualifiers
FEATURES
source
1..281349
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone_lib="DANA-385C13"
/clone_1b="DNA-Arcs BAC library MANN.1"
1..94669
/note="assembly_fragment:00041"
94770..98769
/note="assembly_fragment:02025"
98870..102869
/note="assembly_fragment:02026"
102970..106969
/note="assembly_fragment:02027"
107070..111069
/note="assembly_fragment:02028"
111170..115169
/note="assembly_fragment:02029"
115270..119269
/note="assembly_fragment:02030"
119370..123469
/note="assembly_fragment:02031"
123470..127469
/note="assembly_fragment:02032"
127570..131569
/note="assembly_fragment:02033"
131670..135669
/note="assembly_fragment:02034"
135770..139769
/note="assembly_fragment:02035"
139870..143869
/note="assembly_fragment:02036"
143970..147969
/note="assembly_fragment:02037"
148070..152069
/note="assembly_fragment:02038"
152170..156169
/note="assembly_fragment:02039"
156270..160269
/note="assembly_fragment:02040"
160370..164369
/note="assembly_fragment:02041"
164470..168469
/note="assembly_fragment:02042"
168570..172569
/note="assembly_fragment:02043"
172670..176669
/note="assembly_fragment:02044"
176770..180769
/note="assembly_fragment:02045"
```

```

misc_feature      180870..184869
                  /note="assembly fragment:02054"
misc_feature      189070..193069
                  /note="assembly fragment:02055"
misc_feature      193170..197169
                  /note="assembly fragment:02056"
misc_feature      197270..201269
                  /note="assembly fragment:02057"
misc_feature      201370..205369
                  /note="assembly fragment:02058"

```

```

Query Match      83.2%; Score 20.8; DB 14; Length 281349;
Best Local Similarity 91.7%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY      1 AAAAAAAAACTAGCTGATCT 24
Db      47835 AAAAAAAAACTAGCTGATCT 47858

```

```

RESULT 34
AC083784
LOCUS      AC083784      68644 bp      DNA      linear      HTG 30-SEP-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-481C8 map 2, LOW-PASS SEQUENCE
ACCESSION      AC083784
VERSION      AC083784.1 GI:10440691
KEYWORDS      HTG; HTGS PHASFO.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homidae; Homo.
REFERENCE      1 (bases 1 to 68644)
              Birren,B., Linton,L., Nusbaum,C. and Lander,E.
              Homo sapiens chromosome 2, clone RP11-481C8
              Unpublished
              2 (bases 1 to 68644)
              Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Castle,A.,
              Bouhagalter,B., Brown,A., Burkett,G., Campopiano,A., Castele,A.,
              Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
              Deaellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
              Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
              Graham,L., Grand-pierre,N., Hagos,B., Heaford,A., Horton,L.,
              Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Labrecque,K.,
              Lamasares,R., Landers,T., Leoczky,J., Levine,R., Lieu,C., Liu,G.,
              MacDonald,P., Marguis,N., McCarthy,M., McEwan,P., McKernan,K.,
              McPheeters,R., Meldrum,U., Menus,L., Mihova,T., Mlenga,V.,
              Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
              O'Donnell,P., O'Neill,D., Olivar,T.M., Oliver,C., Riebeck,M., Riley,R.,
              Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
              Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Struass,N., Subramanian,A., Talamas,U., Testaye,S., Theodore,J.,
              Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
              Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J.,
              Zamboni,A. and Zody,M.

```

```

JOURNAL
COMMENT      Submitted (30-SEP-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              All repeats were identified using RepeatMasker:
              Smit,A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

-----Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence submissions@genome.wi.mit.edu
-----Project Information
Center project name: L11230

```

Center clone name: 481_C8

* NOTE: This record contains 84 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

```

1      688: contig of 688 bp in length
      689      788: gap of 100 bp
      789      1516: contig of 728 bp in length
      1517      1616: gap of 100 bp
      1617      2343: contig of 727 bp in length
      2344      2444: gap of 100 bp
      2444      3177: contig of 734 bp in length
      3178      3278: gap of 100 bp
      3278      4001: contig of 723 bp in length
      4001      4101: gap of 100 bp
      4101      4803: contig of 703 bp in length
      4804      4904: gap of 100 bp
      4904      5614: contig of 711 bp in length
      5615      5714: gap of 100 bp
      5715      6432: contig of 718 bp in length
      6433      6532: gap of 100 bp
      6533      7251: contig of 718 bp in length
      7251      7351: gap of 100 bp
      7351      8070: contig of 720 bp in length
      8070      8171: gap of 100 bp
      8171      8876: contig of 706 bp in length
      8877      8976: gap of 100 bp
      8977      9710: contig of 734 bp in length
      9711      9810: gap of 100 bp
      9811      10524: contig of 714 bp in length
      10525      10624: gap of 100 bp
      10625      11353: contig of 729 bp in length
      11354      11453: gap of 100 bp
      11454      12167: contig of 714 bp in length
      12168      12267: gap of 100 bp
      12268      12995: contig of 728 bp in length
      12996      13095: gap of 100 bp
      13096      13812: contig of 717 bp in length
      13813      13912: gap of 100 bp
      13913      14607: contig of 695 bp in length
      14608      14707: gap of 100 bp
      14708      15426: contig of 719 bp in length
      15427      15526: gap of 100 bp
      15527      16253: contig of 727 bp in length
      16254      16353: gap of 100 bp
      16354      17064: contig of 721 bp in length
      17065      17164: gap of 100 bp
      17165      17895: contig of 731 bp in length
      17896      17995: gap of 100 bp
      17996      18717: contig of 722 bp in length
      18718      18817: gap of 100 bp
      18818      19540: contig of 723 bp in length
      19541      19640: gap of 100 bp
      19641      20344: contig of 704 bp in length
      20345      20444: gap of 100 bp
      20445      21123: contig of 679 bp in length
      21124      21223: gap of 100 bp
      21224      21947: contig of 724 bp in length
      21948      22047: gap of 100 bp
      22048      22763: contig of 716 bp in length
      22764      22863: gap of 100 bp
      22864      23581: contig of 718 bp in length
      23582      24367: gap of 100 bp
      24368      24467: contig of 686 bp in length
      24467      24467: gap of 100 bp

```

```

* 24468 25195: contig of 728 bp in length
* 25196 25295: gap of 100 bp
* 25296 26027: contig of 732 bp in length
* 26028 26127: gap of 100 bp
* 26128 26588: contig of 731 bp in length
* 26859 26959: gap of 100 bp
* 27003 27703: contig of 745 bp in length
* 27704 27804: gap of 100 bp
* 28525 28525: contig of 722 bp in length
* 28526 28625: gap of 100 bp
* 28626 29350: contig of 725 bp in length
* 29351 29450: gap of 100 bp
* 29451 30165: contig of 715 bp in length
* 30166 30265: gap of 100 bp
* 30266 30994: contig of 729 bp in length
* 30995 31819: contig of 725 bp in length
* 31820 31919: gap of 100 bp
* 31920 32636: contig of 717 bp in length
* 32637 32736: gap of 100 bp
* 32737 33450: contig of 714 bp in length
* 33451 33550: gap of 100 bp
* 33551 34286: contig of 736 bp in length
* 34287 34386: gap of 100 bp
* 34387 35095: contig of 710 bp in length
* 35096 35196: gap of 100 bp
* 35197 35930: contig of 734 bp in length
* 35931 36030: gap of 100 bp
* 36031 36755: contig of 725 bp in length
* 36756 36855: gap of 100 bp
* 36856 37573: contig of 718 bp in length
* 37574 37673: gap of 100 bp
* 37674 38377: contig of 704 bp in length
* 38378 38477: gap of 100 bp
* 38478 39174: contig of 697 bp in length
* 39175 39274: gap of 100 bp
* 39275 39981: contig of 707 bp in length
* 39982 40081: gap of 100 bp
* 40082 40805: contig of 724 bp in length
* 40806 40905: gap of 100 bp
* 40906 41639: contig of 734 bp in length
* 41640 41739: gap of 100 bp
* 41740 42445: contig of 706 bp in length
* 42446 42545: gap of 100 bp
* 42546 43271: contig of 726 bp in length
* 43272 43371: gap of 100 bp
* 43372 44094: contig of 723 bp in length
* 44095 44194: gap of 100 bp
* 44195 44902: contig of 708 bp in length
* 44903 45002: gap of 100 bp
* 45003 45730: contig of 728 bp in length
* 45731 45830: gap of 100 bp
* 45831 46559: contig of 729 bp in length
* 46560 46659: gap of 100 bp
* 46660 47385: contig of 726 bp in length
* 47386 47485: gap of 100 bp
* 47486 48203: contig of 718 bp in length
* 48204 48303: gap of 100 bp
* 48304 49023: contig of 720 bp in length
* 49024 49123: gap of 100 bp
* 49124 49834: contig of 711 bp in length
* 49835 49934: gap of 100 bp
* 49935 50653: contig of 719 bp in length
* 50654 50753: gap of 100 bp
* 50754 51481: contig of 728 bp in length
* 51482 51581: gap of 100 bp
* 51582 52315: contig of 734 bp in length
* 52316 52415: gap of 100 bp
* 52416 53130: contig of 715 bp in length
* 53131 53230: gap of 100 bp
* 53231 53948: contig of 718 bp in length
* 53949 54048: gap of 100 bp
* 54049 54760: contig of 712 bp in length

```

```

* 54761 54860: gap of 100 bp
* 54861 55564: contig of 704 bp in length
* 55565 55664: gap of 100 bp
* 55665 56370: contig of 706 bp in length
* 56371 56470: gap of 100 bp

Query Match      81.6%; Score 20.4; DB 14; Length 68644;
Best Local Similarity 95.5%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy      1  AAAAAAAAAAAGTACCTGAT 22
Db      15996  AAAAAAAAAAAGTACCTGAT 16017

RESULT 35
AC083784/c
LOCUS
DEFINITION
Homo sapiens chromosome 2 clone RP11-481c8 map 2, LOW-PASS SEQUENCE
SAMPLING.
AC083784
AC083784.1 GI:10440691
HTG: HTGS PHASED.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 68644)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-481c8
Unpublished
2 (bases 1 to 68644)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguski,K., L.,
Bouckgeert,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,I., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Kartas,A., Larocque,K.,
Lamaras,R., Lander,E., Lehotzky,J., Levine,R., Liu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,C., Peterson,K.,
Pierre,N., Pisanu,C., Pollara,V., Raymond,C., Rebeck,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viet,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.
Direct Submision
Submitted (30-SEP-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11230
Center clone name: 481_C_8
-----
* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for

```

```
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
688: contig of 688 bp in length
689
789: gap of 100 bp
789
1516: contig of 728 bp in length
1517
1616: gap of 100 bp
1617
2343: contig of 727 bp in length
2344
2443: gap of 100 bp
2444
3177: contig of 734 bp in length
3178
3277: gap of 100 bp
3278
3278
4100: contig of 723 bp in length
4001
4803: contig of 703 bp in length
4101
4803: gap of 100 bp
4804
4903: gap of 100 bp
4904
5614: contig of 711 bp in length
5714: gap of 100 bp
5615
5715
6432: contig of 718 bp in length
6433
6532: gap of 100 bp
6533
7250: contig of 718 bp in length
7251
7350: gap of 100 bp
7351
8070: contig of 720 bp in length
8071
8170: gap of 100 bp
8171
8876: contig of 706 bp in length
8877
8976: gap of 100 bp
8977
8977
10624: gap of 100 bp
10624
10625
11353: contig of 729 bp in length
11354
11453: gap of 100 bp
11454
12167: contig of 714 bp in length
12168
12267: gap of 100 bp
12268
12995: contig of 728 bp in length
12996
13095: gap of 100 bp
13096
13812: contig of 717 bp in length
13813
13912: gap of 100 bp
13913
14607: contig of 695 bp in length
14608
14707: gap of 100 bp
14708
15426: contig of 719 bp in length
15427
15526: gap of 100 bp
15527
16253: contig of 727 bp in length
16254
16353: gap of 100 bp
16354
17064: contig of 711 bp in length
17065
17164: gap of 100 bp
17165
17895: contig of 731 bp in length
17896
17995: gap of 100 bp
17996
18717: contig of 722 bp in length
18718
18817: gap of 100 bp
18819
19540: contig of 723 bp in length
19541
19640: gap of 100 bp
19641
20344: contig of 704 bp in length
20345
20444: gap of 100 bp
20445
21123: contig of 679 bp in length
21124
21223: gap of 100 bp
21224
21947: contig of 724 bp in length
21948
22047: gap of 100 bp
22049
22763: contig of 716 bp in length
22764
22863: gap of 100 bp
22864
23581: contig of 718 bp in length
23582
23681: gap of 100 bp
23682
24367: contig of 686 bp in length
24368
24467: gap of 100 bp
24469
25195: contig of 728 bp in length
25196
25295: gap of 100 bp
25296
26027: contig of 732 bp in length
26028
26127: gap of 100 bp
26128
26858: contig of 731 bp in length
26859
26959
27703: contig of 745 bp in length
27704

27803: gap of 100 bp
27804
28525: contig of 722 bp in length
28526
28625: gap of 100 bp
28626
29350: contig of 725 bp in length
29351
29450: gap of 100 bp
29451
30165: contig of 715 bp in length
30166
30265: gap of 100 bp
30266
30994: contig of 729 bp in length
30995
31094: gap of 100 bp
31095
31819: contig of 725 bp in length
31820
31919: gap of 100 bp
31920
32636: contig of 717 bp in length
32637
32736: gap of 100 bp
32737
33450: contig of 714 bp in length
33451
33550: gap of 100 bp
33551
34286: contig of 736 bp in length
34287
34386: gap of 100 bp
34387
35096: contig of 710 bp in length
35097
35196: gap of 100 bp
35197
35930: contig of 734 bp in length
35931
36030: gap of 100 bp
36031
36755: contig of 725 bp in length
36756
36855: gap of 100 bp
36856
37573: contig of 718 bp in length
37574
37673: gap of 100 bp
37674
38377: contig of 704 bp in length
38378
38477: gap of 100 bp
38478
39174: contig of 697 bp in length
39175
39274: gap of 100 bp
39275
39981: contig of 707 bp in length
39982
40081: gap of 100 bp
40082
40805: contig of 724 bp in length
40806
40905: gap of 100 bp
40906
41639: contig of 734 bp in length
41640
41739: gap of 100 bp
41740
42445: contig of 706 bp in length
42446
42545: gap of 100 bp
42546
43271: contig of 726 bp in length
43272
43371: gap of 100 bp
43372
44094: contig of 723 bp in length
44095
44194: gap of 100 bp
44195
44902: contig of 708 bp in length
44903
45002: gap of 100 bp
45003
45730: contig of 728 bp in length
45731
45830: gap of 100 bp
45831
46559: contig of 729 bp in length
46560
46659: gap of 100 bp
46660
47385: contig of 726 bp in length
47386
47485: gap of 100 bp
47486
48203: contig of 718 bp in length
48204
48303: gap of 100 bp
48304
49023: contig of 720 bp in length
49024
49123: gap of 100 bp
49124
49834: contig of 711 bp in length
49835
49934: gap of 100 bp
49936
50653: contig of 719 bp in length
50654
50753: gap of 100 bp
50754
51481: contig of 728 bp in length
51482
51581: gap of 100 bp
51582
52315: contig of 734 bp in length
52316
52415: gap of 100 bp
52416
53130: contig of 715 bp in length
53131
53230: gap of 100 bp
53231
53948: contig of 718 bp in length
53949
54048: gap of 100 bp
54049
54760: contig of 712 bp in length
54761
54860: gap of 100 bp
54861
55564: contig of 704 bp in length
55565
55664: gap of 100 bp
55670: contig of 706 bp in length
55671
56470: gap of 100 bp
56471

81.6%, Score 20.4; DB 14; Length 68644;
```

Query Match

Best Local Similarity 95.5%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGAT 22
Db 45237 AAAAAAAAACTATAGCTTGAT 45216

RESULT 36
AE014175_2/c
WPCOMMENT

Sequence split into 4 fragments LOCUS AE014175 Accession AE014175

Fragment Name	Begin	End
AE014175_0	1	110000
AE014175_1	100001	210000
AE014175_2	200001	310000
AE014175_3	300001	404829

Continuation (3 of 4) of AE014175 from base 200001 (AE014175 Mus musculus piebald deleti

Query Match 81.6%; Score 20.4; DB 9; Length 110000;
Best Local Similarity 95.5%; Pred. No. 7.2e+02;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGAT 22
Db 55107 AAAAAAAAACTATAGCTTGAT 55086

RESULT 37
AC096773 136403 bp DNA linear PRI 01-MAR-2002
LOCUS
DEFINITION Homo sapiens BAC clone RP11-810P8 from 4, complete sequence.
AC096773 AC022697
VERSION AC096773.3 GI:15920158
KEYWORDS
HTG.
SOURCE Homo sapiens (human)

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 136403)
TITLE Sulston, J.E. and Waterston, R.
JOURNAL Toward a complete human genome sequence
PUBMED Genome Res. 8 (11), 1097-1108 (1998)
9847074

REFERENCE
AUTHORS 2 (bases 1 to 136403)
TITLE Grewal, N. and Kozlowski, A.
JOURNAL The sequence of Homo sapiens BAC clone RP11-810P8
Unpublished (2001)
3 (bases 1 to 136403)

REFERENCE
AUTHORS Waterston, R.H.
JOURNAL Direct Submission
TITLE Submitted (25-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 136403)

REFERENCE
AUTHORS Waterston, R.H.
JOURNAL Direct Submission
TITLE Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 136403)

REFERENCE
AUTHORS Waterston, R.
JOURNAL Direct Submission
TITLE Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gi:15778810.

COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_NH0810P08
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at <http://www.chori.org>
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-123G5; the clone sequenced to the right is RP11-125018, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-810P8; actual end is at base position 33102 of RP11-125018.

FEATURES
The sequence of AC022697 has been incorporated into AC096773.

source location/Qualifiers

1..136403 "Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-810P8"
/clone_11b="RPCT-11"
1..2375
/rpt_family="L1"
repeat_region
2396..2685
/rpt_family="Alu"
repeat_region
2659..2726
/rpt_family="(TAAA)n"
repeat_region
2731..5308
/rpt_family="L1"
repeat_region
5700..5725
/rpt_family="AT-rich"
5748..6253
/note="similar to EST B1495543 (NID:gi15334887)"
5856..6253
/note="similar to EST B1495544 (NID:gi15334888)"
6056..6212
/rpt_family="L1"
repeat_region
6220..6329
/rpt_family="L1"
repeat_region
6315..6337
/rpt_family="(CAAA)n"
7984..8021
/rpt_family="(CA)n"
8654..9018
/rpt_family="L2"

```

repeat_region      9409. .9472
                    /rpt_family="(CA)n"
repeat_region      10095. .10316
                    /rpt_family="MIR"
repeat_region      10312. .10342
                    /rpt_family="AT_rich"
repeat_region      12732. .12931
                    /rpt_family="L2"
repeat_region      14041. .14083
                    /rpt_family="L2"
repeat_region      14322. .14382
                    /rpt_family="MIR"
repeat_region      14558. .14664
                    /rpt_family="MIR"
repeat_region      16158. .16333
                    /rpt_family="MERS3"
repeat_region      16242. .16274
                    /rpt_family="AT_rich"
repeat_region      17086. .17177
                    /rpt_family="(TA)n"
repeat_region      17619. .18051
                    /rpt_family="L1"
repeat_region      18616. .19721
                    /rpt_family="L1"
repeat_region      19746. .20068
                    /rpt_family="L1"
repeat_region      20076. .21378
                    /rpt_family="L1"
repeat_region      21305. .21332
                    /rpt_family="AT_rich"
repeat_region      21379. .21683
                    /rpt_family="Alu"
repeat_region      21684. .22778
                    /rpt_family="L1"
repeat_region      23041. .23274
                    /rpt_family="MERS2_type"
repeat_region      23464. .23967
                    /rpt_family="ERVK"
repeat_region      24426. .24872
                    /rpt_family="MERS2_type"
repeat_region      24569. .24598
                    /rpt_family="AT_rich"
repeat_region      24863. .24887
                    /rpt_family="(T)n"
repeat_region      30965. .31075
                    /rpt_family="ERV1"
repeat_region      31110. .31198
                    /rpt_family="MERS2_type"
repeat_region      33303. .33396
                    /rpt_family="L2"
repeat_region      34516. .34540
                    /rpt_family="AT_rich"
repeat_region      34829. .34863
                    /rpt_family="(CAA)n"
repeat_region      35255. .35598
                    /rpt_family="MALR"
repeat_region      36050. .36757
                    /rpt_family="MALR"
repeat_region      37272. .37347
                    /rpt_family="T-rich"
repeat_region      37409. .37430
                    /rpt_family="AT_rich"
repeat_region      37466. .37576
                    /rpt_family="L1"
repeat_region      40116. .40191
                    /rpt_family="L1"
repeat_region      40192. .40778
                    /rpt_family="L1"
repeat_region      40764. .40787
                    /rpt_family="AT_rich"
repeat_region      40997. .41125

```

```

repeat_region      /rpt_family="L1"
repeat_region      41132. .41441
                    /rpt_family="Alu"
repeat_region      41415. .41441
                    /rpt_family="(A)n"
repeat_region      42140. .42169
                    /rpt_family="AT_rich"
misc_feature       45895. .46271
                    /note="similar to EST AW020477 (NID:gl5330161) df10d04.y1"
misc_feature       45895. .46266
                    /note="similar to EST B1491501 (NID:gl5330845)"

Query Match      81.6%; Score 20.4; DB 8; Length 136403;
Best Local Similarity 95.5%; Pred. No. 6.8e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1
1 AAAAAAAAAAAGTATGCTTGAT 22
DB 114236 AAAAAAAAAAAGTATGCTTGAT 114257

RESULT 38
AC021878      161391 bp DNA linear PRI 13-MAY-2005
LOCUS      Homo sapiens BAC clone RP11-18N21 from 4, complete sequence.
DEFINITION
AC021878
ACCESSION
VERSION      AC021878.4 GI:11120931
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 161391)
Kang,K., Maupin,R. and Parker,C.
The sequence of Homo sapiens BAC clone RP11-18N21
Unpublished (2001)
2 (bases 1 to 161391)
Waterston,R.H.
Direct Submisison
Submitted (21-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 161391)
Waterston,R.H.
Direct Submisison
Submitted (08-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 161391)
Waterston,R.
Direct Submisison
Submitted (07-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 161391)
Wilson,R.K.
Direct Submisison
Submitted (13-MAY-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Nov 8, 2000 this sequence version replaced gi:8569722.

COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
-----
Project name: H_NH0018N21

```

NOTICE:

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-18N21; actual end is at base position 161391 of RP11-18N21.

FEATURES

SOURCE

```
1. 161391
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="4"
   /clone_1db="RPCI-11"
   /clone_1id="RP11-18N21"
   /note="Cpg island (4GC=62.4, o/e=0.89, #CpGs=24)"
   misc_feature
   140332..140605
   /note="Cpg_island (4GC=62.0, o/e=0.85, #CpGs=23)"
```

ORIGIN

```
Query Match      81.6%  Score 20.4  DB 8:  Length 161391;
                  45.5%  P-val: 6.5e-02;
                  Mismatch: 1;  Indels 0;  Gaps 0;
```

QY

```
1 AAAAAAAAACTATGCTTGAT 22
  |||||
```

```
Db      143793 AAAAAAAAACTATGCTTGAT 143814
```

RESULT 39

AC022999

LOCUS

AC022999 Homo sapiens chromosome 19 clone RP11-806X22 map 19, WORKING DRAFT

DEFINITION

SEQUENCE, 15 unordered pieces.

AC022999

AC022999.3 GI:10198518

HTG: HTGS PHASE1: HTGS_DRAFT.

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE

1 (bases 1 to 164988)

Birtten, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 19, clone RP11-806X22

JOURNAL

AUTHORS

Birtten, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, D., Barua, N., Beckerly, R., Beda, F., Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., D'Arrellano, K., Dewar, K., Domino, M., Doyle, M., Fenebor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,

TITLE

JOURNAL

COMMENT

Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lebowitz, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McEwan, P., McGuck, A., McKernan, K., McPherson, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, W.

Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:9164582.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L6587

Center clone name: 806_K_22

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 157821 bases at least Q40

Consensus quality: 161489 bases at least Q30

Consensus quality: 162773 bases at least Q20

Insert size: 170000; agarose-gel

Insert size: 163588; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-gel

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 11415: contig of 11415 bp in length
11416 11515: gap of 100 bp
11516 15092: contig of 3577 bp in length
15093 15193: gap of 100 bp
15193 22042: contig of 6850 bp in length
22043 22143: gap of 100 bp
22143 26809: contig of 4667 bp in length
26810 26909: gap of 100 bp
26909 33948: contig of 7039 bp in length
33949 34048: gap of 100 bp
34049 41393: contig of 7345 bp in length
41394 41493: gap of 100 bp
41494 47129: contig of 5636 bp in length
47130 47230: gap of 100 bp
47230 55490: contig of 8261 bp in length
55491 55591: gap of 100 bp
55591 63721: contig of 8131 bp in length
63722 63821: gap of 100 bp
63822 72604: contig of 8783 bp in length
72605 72704: gap of 100 bp
72705 84834: contig of 12130 bp in length
84835 84934: gap of 100 bp
84935 98028: contig of 13094 bp in length
98029 98128: gap of 100 bp
98129 124081: contig of 25953 bp in length
124082 124181: gap of 100 bp
124182 146875: contig of 22694 bp in length
```

```

FEATURES
 * 146876 146975: gap of 100 bp
 * 146976 164988: contig of 18013 bp in length.
Location/Qualifiers
  source
    1. 164988
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="19"

misc_feature
  1. 11415
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:left"
  11416..11515
  /estimated_length=100
  /note="assembly_fragment"
  15093..15192
  /estimated_length=100
  15193..22042
  /note="assembly_fragment"
  22043..22142
  /estimated_length=100
  22143..26809
  /note="assembly_fragment"
  26810..26909
  /estimated_length=100
  26910..33948
  /note="assembly_fragment"
  33949..34048
  /estimated_length=100
  34049..41393
  /note="assembly_fragment"
  41394..41493
  /estimated_length=100
  41494..47129
  /note="assembly_fragment"
  47130..47229
  /estimated_length=100
  47230..55490
  /note="assembly_fragment"
  55491..55590
  /estimated_length=100
  55591..63721
  /note="assembly_fragment"
  63722..63821
  /estimated_length=100
  63822..72604
  /note="assembly_fragment"
  72605..72704
  /estimated_length=100
  72705..84834
  /note="assembly_fragment"
  84835..84934
  /estimated_length=100
  84935..98028
  /note="assembly_fragment"
  98029..98128
  /estimated_length=100
  98129..124081
  /note="assembly_fragment"
  124082..124181
  /estimated_length=100
  124182..146875
  /note="assembly_fragment"
  146876..146975
  /estimated_length=100
  146976..164988
  /note="assembly_fragment"
  clone_end:T7
  vector_side:right"

```

```

ORIGIN
Query Match
Best Local Similarity 81.6%; Score 20.4; DB 14; Length 164988;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGAT 22
Db 14709 AAAAAAAAACTATAGCTTTAT 14730

RESULT 40
AC008529
LOCUS AC008529 169557 bp DNA linear PRI 26-JAN-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-475P15, complete sequence.
ACCESSION AC008529
VERSION AC008529.4 GI:12545282
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE 1 (bases 1 to 169557)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169557)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94596, USA
3 (bases 1 to 169557)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jan 26, 2001 this sequence version replaced gi:7708931.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
www.jgi.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES
  source
    1. 169557
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="5"
    /clone="CTC-475P15"

ORIGIN
Query Match
Best Local Similarity 81.6%; Score 20.4; DB 8; Length 169557;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGAT 22
Db 94720 AAAAAAAAACTATAGCTTTAT 94741

RESULT 41
CT025554
LOCUS CT025554 201412 bp DNA linear HTG 04-AUG-2005
DEFINITION Mus musculus chromosome 14 clone RP23-469A12, *** SEQUENCING IN
PROGRES ***. 6 unordered pieces.
ACCESSION CT025554
VERSION CT025554.2 GI:71833838
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS
TITLE
JOURNAL

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 201412)

Direct Submission

Submitted (03-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Aug 4, 2005 this sequence version replaced gi:71794513.
----- Genome Center

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: vegas@sanger.ac.uk

Project Information

Center project name: BM469A12

Summary Statistics

Assembly program: XGAP; version 4.5

Chemistry: Dye-terminator; 10% of reads

Consensus quality: 199579 bases at least Q40

Consensus quality: 200175 bases at least Q30

Consensus quality: 200603 bases at least Q20

Insert size: 200912; sum-of-contigs

Insert size: 201377; 2.9% error; agarose-fp

Quality coverage: 6.90x in Q20 bases; sum-of-contigs Quality
coverage: 7.01x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 30988: contig of 30988 bp in length
* 30989 31088: gap of 100 bp
* 31089 33594: contig of 2506 bp in length
* 33595 33694: gap of 100 bp
* 33695 63502: contig of 29808 bp in length
* 63503 63602: gap of 100 bp
* 63603 141278: contig of 77676 bp in length
* 141279 141378: gap of 100 bp
* 141379 184269: contig of 42891 bp in length
* 184270 184369: gap of 100 bp
* 184370 201412: contig of 17043 bp in length.

----- Annotation/Qualifiers

1 30988

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="14"

/clone="RP23-469A12"

/clone_id="RP23-469A12"

1. 30988

/note="assembly_fragments:00216"

clone_end:SP6

vector_side:left

31089..33594

/note="assembly_fragments:00014"

fragment_chain:1

33695..63502

/note="assembly_fragments:00483"

fragment_chain:1

63603..141278

/note="assembly_fragments:01263"

fragment_chain:1

141379..184269

/note="assembly_fragments:00772"

fragment_chain:1

184370..201412

/note="assembly_fragments:00042"

clone_end:T7

ORIGIN

vector_side:right"

Query Match

Best Local Similarity 81.6%; Score 20.4; DB 14; Length 201412;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AAAAAAAAACTAGCTGAT 22

29945 AAAAAAAAACTAGCTGAT 29966

RESULT 42

AC114410/c

LOCUS

DEFINITION

AC114410

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 201783)

Birren, B., Nussbaum, C. and Lander, E.

Mus musculus, clone RP23-151K8

Unpublished

2 (bases 1 to 201783)

Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Boukigalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., Dettellano, K., Dewar, K., Diaz, J. S., Dodge, S.,

Faro, S., Ford, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Ginde, S., Gordon, L., Hulse, W., Iliev, I., Johnson, R., Jones, C.,

Kamet, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,

Lander, T., Lehotzky, J., Levine, R., Liu, G., Maclean, C.,

Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M.,

McEwan, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T.,

Melena, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,

Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupbach, R., Seaman, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Straus, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trifilio, J., Vasilleff, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (08-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 201783)

Birren, B., Nussbaum, C. and Lander, E.

Direct Submission

Submitted (15-AUG-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 201783)

Birren, B., Nussbaum, C. and Lander, E.

Direct Submission

Submitted (09-SEP-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 9, 2003 this sequence version replaced gi:33667216.

All repeats were identified using RepeatMasker:

Smith, A. F. A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

```

----- Project Information
Center project name: L23439
Center clone name: 151_K_8

FEATURES
-----
source
Location/Qualifiers
1. .201763
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-151K8"
/clone_lib="RP23-151K8"
misc_feature
1. .7101
/note="qgs end extension
clone end:SP6"
repeat_region
complement(90..201)
/rpt_family="PBID10"
repeat_region
443..539
/rpt_family="B1F"
repeat_region
546..574
/rpt_family="TCCA)n"
/rpt_family="C9C"
/rpt_family="C9C"

/rpt_family="A)n"
1436..1618
/rpt_family="B2_Mm2"
2164..2185
/rpt_family="AT_rich"
complement(3339..3466)
/rpt_family="B1F"
complement(4368..4448)
/rpt_family="L1_MM"
7102..7107
/note="clone boundary
clone end:SP6
site:ECORI"
9053..9152
/rpt_family="TA)n"
12029..12129
/rpt_family="MTE"
12236..12345
/rpt_family="PBID7"
12752..12867
/rpt_family="(TC)n"
complement(13407..13609)
/rpt_family="B3"
14127..14167
/rpt_family="(TA)n"
14167..14193
/rpt_family="(CA)n"
14235..14272
/rpt_family="(CA)n"
14521..14558
/rpt_family="(TTA)n"
14935..14963
/rpt_family="AT_rich"
15589..15629
/rpt_family="(TTTG)n"
complement(15632..15765)
/rpt_family="B1_MM"
15919..16063
/rpt_family="B1_MM"
16068..16090
/rpt_family="(CAAC)n"
16143..16166
/rpt_family="(CAA)n"
complement(17349..17514)
/rpt_family="RSINE1"
18604..18631
/rpt_family="(A)n"
18696..18735
/rpt_family="(A)n"
19686..19753

```

```

repeat_region /rpt_family="CT-rich"
22505..22598
/rpt_family="(TTTC)n"
repeat_region /rpt_family="CT-rich"
22891..22945
/rpt_family="(CT-rich"
complement(24470..24595)
/rpt_family="RSINE1"
25058..25105
/rpt_family="(TATA)n"
26176..26465
/rpt_family="B4A"
26473..26541
/rpt_family="(CA)n"
26917..26939
/rpt_family="(CAAA)n"
28766..28877
/rpt_family="PBID9"
28878..28909
/rpt_family="(CAA)n"
29243..29455
/rpt_family="B4"
30293..30316
/rpt_family="AT_rich"
complement(30557..31009)
/rpt_family="L1"
complement(31159..31213)
/rpt_family="L1"
31215..32472
/rpt_family="Lx4"
complement(33069..33221)
/rpt_family="L1_MM"
33438..33470
/rpt_family="(CAA)n"
complement(36123..36255)
/rpt_family="B1F"
37393..37422
/rpt_family="AT_rich"
40393..40422
/rpt_family="(TA)n"
40692..40712
/rpt_family="AT_rich"
41263..41297
/rpt_family="AT_rich"
42833..42861
/rpt_family="AT_rich"
43372..43402
/rpt_family="AT_rich"
complement(44537..44680)
/rpt_family="B1_MM"
complement(44781..44978)
/rpt_family="URR1A"
44993..45026
/rpt_family="AT_rich"
45864..45939
/rpt_family="CT-rich"
45946..45982
/rpt_family="(CA)n"
complement(47163..47222)
/rpt_family="S8"
47927..48187
/rpt_family="Lx8"
48211..48294
/rpt_family="Lx8"
48808..48990
/rpt_family="B3"
complement(50184..50368)
/rpt_family="B3"
50559..50602

```

Query Match 81.6%; Score 20.4; DB 9; Length 201763;
 Best Local Similarity 95.5%; Pred. No. 6; 1e+02;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTGAT 22
Db 12681 AAAAAAAAACTAGCATGAT 126860

RESULT 43
AX346556/c
LOCUS AX346556 5228 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 1627 from Patent WO0200928.
ACCESSION AX346556
VERSION AX346556.1 GI:18494442
KEYWORDS
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.

REFERENCE
1 Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1627 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
source
1. 5228
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 80.8%; Score 20.2; DB 6; Length 5228;
Best Local Similarity 88.0%; Pred. No. 2e+03; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTGATCTT 25
Db 336 AAAAAAAAACTAGCTTTTCTT 312

RESULT 44
AC014427/c
LOCUS AC014427 6584 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
ACCESSION AC014427
VERSION AC014427.1 GI:6436908
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 6584)
AUTHORS Adams, M. and Venter, J.C.
JOURNAL Direct Submission
COMMENT Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
This sequence was identified as CDM:10210601 by the submitter. For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. 6584
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 80.8%; Score 20.2; DB 14; Length 6584;
Best Local Similarity 88.0%; Pred. No. 1.9e+03; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTGATCTT 25
Db 1074 AAAAAAAAACTAGCTTATCTT 1050

RESULT 45
AX348975/c
LOCUS AX348975 15649 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 433 from Patent WO0202807.
ACCESSION AX348975
VERSION AX348975.1 GI:18615010
KEYWORDS
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.

REFERENCE
1 Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with cell signalling
JOURNAL Patent: WO 0202807-A 433 10-JAN-2002;
Epigenomics AG (DE)
FEATURES
source
1. 15649
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 80.8%; Score 20.2; DB 6; Length 15649;
Best Local Similarity 88.0%; Pred. No. 1.5e+03; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTGATCTT 25
Db 5501 AAAAAAAAACTAGCTTAATCTT 5477

RESULT 46
AY821561/c
LOCUS AY821561 18525 bp DNA linear INV 01-MAY-2005
DEFINITION Manduca sexta chitin synthase 2 gene, complete cds.
ACCESSION AY821561
VERSION AY821561.1 GI:60459511
KEYWORDS
SOURCE Manduca sexta (tobacco hornworm)
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Dictyostea; Sphingidae; Sphingidae; Manduca.
1 (bases 1 to 18525)
AUTHORS Hogenkamp, D.G., Arakane, Y., Zimoch, L., Merzendorfer, H., Kramer, K.J., Beeman, R.W., Kanost, M.R., Specht, C.A. and Muthukrishnan, S.
TITLE Chitin synthase gene in Manduca sexta: characterization of a gut-specific transcript and differential tissue expression of alternatively spliced mRNAs during development
JOURNAL Insect Biochem. Mol. Biol. 35 (6), 529-540 (2005)
PUBMED 15857759
REFERENCE 2 (bases 1 to 18525)
AUTHORS Hogenkamp, D.G., Arakane, Y., Zimoch, L., Merzendorfer, H., Kramer, K.J., Beeman, R.W., Kanost, M.R., Specht, C.A. and Muthukrishnan, S.
JOURNAL Direct Submission
COMMENT Submitted (08-NOV-2004) Biochemistry, Kansas State University, 104 Willard Hall, Manhattan, KS 66506, USA
FEATURES
source
1. 18525
/organism="Manduca sexta"
/mol_type="genomic DNA"
/db_xref="taxon:7130"
Join11. 166,448 .562,3845. .4089,4323. .4489,4699. .5029,5228. .5424,5706. .6015,7146. .7366,7867. .7952,8189. .8340,9380. .9642,9953. .10163,10355. .10555,11115. .11341,11967. .12134,12374. .12679,12960. .13189,13563. .13689,14169. .14321,14879. .15055,15715. .15882,17379. .17610,

CDs

```

17770..17936,18340..18525)
/product="chitin synthase 2"
/join(465..562,3845..4089,4323..4489,4699..5029,5228..5424,
5706..6015,7146..7366,7867..7952,8189..8340,9380..9642,
9953..10163,10355..10555,11115..11341,11967..12134,
12679..12960..13189,13563..13689,14169..14321,
14879..15055,15715..15882,17379..17610,17770..17936,
18340..18477)
/EC_number="2.4.1.16"
/note="MacH2"
/codon_start=1
/product="chitin synthase 2"
/protein_id="AA020092.1"
/db_xref="GI:60459512"
/translation="MAATPGFKLADSESDTEYTPLYDGDIDORTAQTGKN
LREPLPVKSGSMATKMIETSVKIKYLAIVLCAVLSGAVLAKGLPITSLQK
KROITPCNRRLADQOITVHSLERITLMAALVPGVLEBSVRICTFETA
KRPOTPIATFTETLOAIGALVLLILPEIDAVKAMNATCAIPALINIFRD
RMDKPSIKLIDVLAISAQATFVWPMPMPVMTI PVAQVLSGFMENFTDY
NKSVPFTVGLRDNIKRTRYTORLSTVSKIIIPACILISLHMONDNPFTTHAS
KAPTRVYVNVIVVRDRTTIGDVTGIFRDLITSLMVALIOGAAPFCGS
.....
LNLRIKLDLNDUCVVDLKGJLDDSVGSDLAIVKDKPDSITRIQVATMMHETNEE
MIEPLKSIPLDEDOGARVAVOKYIGIVDPDYELSGHFMDDAFISDHSADSOVN
RVKCLVDAVDRASRVHLTVNRLRPKPYPTPYGKLIWTMPGKIKLICHDKSKI
RHKRMSOVWYVYFGHRLMDLPLSVDRKVIATRYVLLADGIDPSPAVTLLIG
GTFPLMIGAINATIGSMNHALLFNLVLTFLVCMTCSTETOLMLNLITCFPAM
VMMFVLSIVLQISODGMLAPSPMAATFGIEFYALALPOEIIICLALISYYITIP
SNYMLIIVISICNLNNVSWGTREVAOKTAKEMDMKAAEAKKMDOSJMKFPGK
SSETSGLSLFSVAGLFRCMCTNPDKHDLILQIANSIEKIKESLALGSESPA
QOCTRSSLSLGRDELATMPYADELSGDI PREERDILNPYVEDNLKXGEYD
LTTAEIEFWKDLIDVLRPIDNKKEOERIKTDIKRLRTPVPAFAMNSLFLVILYF
LOPNDOLIKMPFQODVLAISYDKERNVYLQOEPLMEPIGSLPIVFGPMLIOP
AMLSHSYITTHLSTELHWFPSRPPQMSDBNLERKVEIEARLQKLVNDDLRK
AVEIDNVSRRKTLHLEKARDTKHSYVNLIDANFKRRLTILQSGDPVIVSRSLSGDE
VRRATIRALKTRDLSLAEKRSQLOAAGDATGYMNLSTGAVNDMSGRASTASAYI
NKGYEAPFSDDEPRPRRTVRFRENT"

```

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 18525;
 Best Local Similarity 88.0%; Pred. No. 1.4e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTGATCTT 25
 Db 11025 AAAAAAAAACTAGCTGATTTT 11001

RESULT 47
 AC067765
 LOCUS
 DEFINITION Homo sapiens chromosome 2 clone RP11-78L3 map 2, LOW-PASS SEQUENCE
 SAMPLING.
 AC067765
 VERSION AC067765.1 GI:7651814
 KEYWORDS HTGS PHASE0.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 1 (bases 1 to 65059)
 Birren,B., Lincon,L., Nuebaum,C. and Lander,E.
 Homo sapiens chromosome 2, clone RP11-78L3
 2 (bases 1 to 65059)
 unpublished
 Birren,B., Lincon,L., Nuebaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,

TITLE
 JOURNAL
 COMMENT

Submitted (27-Apr-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smith, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L9971
 Center clone name: 78_L3

* NOTE: This record contains 79 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 731 830: contig of 730 bp in length
 831 830: gap of 100 bp
 1549 831: contig of 718 bp in length
 1549 831: gap of 100 bp
 1649 831: contig of 729 bp in length
 2378 831: gap of 100 bp
 2478 831: contig of 726 bp in length
 3204 831: gap of 100 bp
 3304 831: contig of 733 bp in length
 4037 831: gap of 100 bp
 4137 831: contig of 704 bp in length
 4841 831: gap of 100 bp
 4941 831: contig of 737 bp in length
 5678 831: gap of 100 bp
 5778 831: contig of 740 bp in length
 6518 831: gap of 100 bp
 6618 831: contig of 721 bp in length
 7339 831: gap of 100 bp
 7439 831: contig of 751 bp in length
 8190 831: gap of 100 bp
 8290 831: contig of 721 bp in length
 9011 831: gap of 100 bp
 9111 831: contig of 727 bp in length
 9838 831: gap of 100 bp
 9938 831: contig of 716 bp in length
 10653 831: gap of 100 bp
 10654 831: contig of 681 bp in length
 11435 831: gap of 100 bp
 12222 831: contig of 688 bp in length

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dogge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,M., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,C., Jones,C., Kenn,L., Karatas,A.,
 Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoscky,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McNetters,R.,
 Melgrim,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talmac,J.,
 Testaye,S., Theodore,J., Tittel,A., Travers,M., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```

* 12223 12322: gap of 100 bp
* 12323 13062: contig of 740 bp in length
* 13063 13162: gap of 100 bp
* 13163 13892: contig of 730 bp in length
* 13893 13992: gap of 100 bp
* 13993 14723: contig of 731 bp in length
* 14724 14823: gap of 100 bp
* 14824 15560: contig of 737 bp in length
* 15561 15660: gap of 100 bp
* 15661 16372: contig of 712 bp in length
* 16373 16472: gap of 100 bp
* 16473 17212: contig of 740 bp in length
* 17213 17312: gap of 100 bp
* 17313 18055: contig of 744 bp in length
* 18057 18156: gap of 100 bp
* 18157 18894: contig of 738 bp in length
* 18895 18994: gap of 100 bp
* 18995 19731: contig of 737 bp in length
* 19732 19831: gap of 100 bp
* 19832 20569: contig of 738 bp in length
* 20570 20669: gap of 100 bp
* 20670 21408: contig of 739 bp in length
* 21409 21508: gap of 100 bp
* 21509 22252: contig of 744 bp in length
* 22253 22352: gap of 100 bp
* 22353 23088: contig of 737 bp in length
* 23090 23189: gap of 100 bp
* 23190 23945: contig of 756 bp in length
* 23946 24045: gap of 100 bp
* 24046 24778: contig of 733 bp in length
* 24779 24878: gap of 100 bp
* 24879 25609: contig of 731 bp in length
* 25610 25710: gap of 100 bp
* 25710 26428: contig of 719 bp in length
* 26429 26528: gap of 100 bp
* 26529 27236: contig of 708 bp in length
* 27237 27336: gap of 100 bp
* 27337 28033: contig of 697 bp in length
* 28034 28133: gap of 100 bp
* 28134 28873: contig of 740 bp in length
* 28874 28973: gap of 100 bp
* 28974 29708: contig of 735 bp in length
* 29709 29808: gap of 100 bp
* 29809 30544: contig of 736 bp in length
* 30545 30644: gap of 100 bp
* 30645 31366: contig of 722 bp in length
* 31367 31466: gap of 100 bp
* 31467 32206: contig of 740 bp in length
* 32207 32306: gap of 100 bp
* 32307 33029: contig of 723 bp in length
* 33030 33129: gap of 100 bp
* 33130 33871: contig of 742 bp in length
* 33872 33971: gap of 100 bp
* 33972 34703: contig of 732 bp in length
* 34704 34803: gap of 100 bp
* 34804 35528: contig of 726 bp in length
* 35529 35628: gap of 100 bp
* 35630 36308: contig of 679 bp in length
* 36309 36408: gap of 100 bp
* 36409 37131: contig of 723 bp in length
* 37132 37231: gap of 100 bp
* 37232 37920: contig of 689 bp in length
* 37921 38020: gap of 100 bp
* 38021 38745: contig of 725 bp in length
* 38746 38845: gap of 100 bp
* 38846 39482: contig of 637 bp in length
* 39483 39582: gap of 100 bp
* 39583 40315: contig of 733 bp in length
* 40316 40415: gap of 100 bp
* 40416 41177: contig of 762 bp in length
* 41178 41277: gap of 100 bp
* 41278 42012: contig of 735 bp in length
* 42013 42112: gap of 100 bp

```

```

* 42113 42830: contig of 718 bp in length
* 42831 42930: gap of 100 bp
* 42931 43631: contig of 701 bp in length
* 43632 43731: gap of 100 bp
* 43732 44457: contig of 726 bp in length
* 44458 44557: gap of 100 bp
* 44558 45244: contig of 687 bp in length
* 45245 45344: gap of 100 bp
* 45345 46072: contig of 728 bp in length
* 46073 46172: gap of 100 bp
* 46173 46913: contig of 741 bp in length
* 46914 47013: gap of 100 bp
* 47014 47743: contig of 730 bp in length
* 47744 47843: gap of 100 bp
* 47844 48574: contig of 731 bp in length
* 48575 48674: gap of 100 bp
* 48675 49402: contig of 728 bp in length
* 49403 49503: gap of 100 bp
* 49503 50238: contig of 736 bp in length
* 50239 50338: gap of 100 bp
* 50339 51061: contig of 723 bp in length
* 51062 51161: gap of 100 bp
* 51162 51864: contig of 703 bp in length
* 51865 51964: gap of 100 bp
* 51965 52692: contig of 728 bp in length
* 52693 52792: gap of 100 bp
* 52793 53505: contig of 713 bp in length
* 53506 53605: gap of 100 bp
* 53606 54448: contig of 743 bp in length
* 54449 54449: gap of 100 bp
* 54449 55169: contig of 721 bp in length
* 55170 55269: gap of 100 bp
* 55270 55997: contig of 728 bp in length
* 55998 56097: gap of 100 bp
* 56098 56833: contig of 736 bp in length

```

```

Query Match      80.8%; Score 20.2; DB 14; Length 65059;
Best Local Similarity 88.0%; Pred. No. 9.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAACTAGCTGATCTT 25
Db 38641 AAAAAAAAACTTTCTGATCTT 38665

```

RESULT 48

AC100199 Mus musculus clone RP23-58K8, LOW-PASS SEQUENCE SAMPLING.

AC100199 AC100199.1 GI:17047565
 VERSION
 HTG; HTGS PHASE0.

KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 1 (bases 1 to 65091)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Brown, A., Camarato, J., Campolano, A., Chang, J., Chazaro, B.,
 Choquel, Y., Colangelo, M., Collins, S., Collins, S., Fero, S.,
 Cooke, P., DeRubeis, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyne, S.,
 Ginde, S., Goid, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hages, B., Heath, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Lander, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,

2 (bases 1 to 65091)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarato, J., Campolano, A., Chang, J., Chazaro, B.,
 Choquel, Y., Colangelo, M., Collins, S., Collins, S., Fero, S.,
 Cooke, P., DeRubeis, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyne, S.,
 Ginde, S., Goid, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hages, B., Heath, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Lander, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,

Unpublished
 1 (bases 1 to 65091)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarato, J., Campolano, A., Chang, J., Chazaro, B.,
 Choquel, Y., Colangelo, M., Collins, S., Collins, S., Fero, S.,
 Cooke, P., DeRubeis, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyne, S.,
 Ginde, S., Goid, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hages, B., Heath, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Lander, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,

TITLE
JOURNAL
COMMENT

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Menas, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Poliera, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rhee, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnapack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Struss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wymen, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smith, F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L14319
Center clone name: 5g_K_8

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 701: contig of 701 bp in length
* 702 801: gap of 100 bp
* 802 1518: contig of 717 bp in length
* 1519 1618: gap of 100 bp
* 1619 2334: contig of 716 bp in length
* 2335 2434: gap of 100 bp
* 2435 3151: contig of 717 bp in length
* 3152 3251: gap of 100 bp
* 3252 3968: contig of 717 bp in length
* 3969 4068: gap of 100 bp
* 4069 4735: contig of 667 bp in length
* 4736 4835: gap of 100 bp
* 4836 5556: contig of 721 bp in length
* 5557 5656: gap of 100 bp
* 5657 6373: contig of 717 bp in length
* 6374 6473: gap of 100 bp
* 6474 7207: contig of 734 bp in length
* 7208 7307: gap of 100 bp
* 7308 8026: contig of 719 bp in length
* 8027 8126: gap of 100 bp
* 8127 8840: contig of 714 bp in length
* 8841 8940: gap of 100 bp
* 8941 9651: contig of 711 bp in length
* 9652 9751: gap of 100 bp
* 9752 10450: contig of 699 bp in length
* 10451 10550: gap of 100 bp
* 10551 11261: contig of 711 bp in length
* 11262 11361: gap of 100 bp
* 11362 12077: contig of 716 bp in length
* 12078 12177: gap of 100 bp
* 12178 12876: contig of 699 bp in length
* 12877 12976: gap of 100 bp
* 12978 13667: contig of 691 bp in length
* 13668 13767: gap of 100 bp
* 13768 14475: contig of 708 bp in length
* 14476 14575: gap of 100 bp

* 14576 15292: contig of 717 bp in length
* 15293 15392: gap of 100 bp
* 15393 16095: contig of 703 bp in length
* 16096 16195: gap of 100 bp
* 16196 16316: contig of 721 bp in length
* 16317 17016: gap of 100 bp
* 17017 17728: contig of 712 bp in length
* 17729 17828: gap of 100 bp
* 17829 18545: contig of 717 bp in length
* 18546 18645: gap of 100 bp
* 18646 19378: contig of 733 bp in length
* 19379 19478: gap of 100 bp
* 19479 20212: contig of 734 bp in length
* 20213 20312: gap of 100 bp
* 20313 21046: contig of 734 bp in length
* 21047 21146: gap of 100 bp
* 21147 21864: contig of 718 bp in length
* 21865 21964: gap of 100 bp
* 21965 22681: contig of 717 bp in length
* 22682 22781: gap of 100 bp
* 22782 23495: contig of 714 bp in length
* 23496 23595: gap of 100 bp
* 23596 24312: contig of 717 bp in length
* 24313 24412: gap of 100 bp
* 24413 25135: contig of 723 bp in length
* 25136 25235: gap of 100 bp
* 25236 25953: contig of 718 bp in length
* 25954 26053: gap of 100 bp
* 26054 26764: contig of 711 bp in length
* 26765 26864: gap of 100 bp
* 26865 27594: contig of 730 bp in length
* 27595 27694: gap of 100 bp
* 27695 28413: contig of 719 bp in length
* 28414 28513: gap of 100 bp
* 28514 29231: contig of 718 bp in length
* 29232 29331: gap of 100 bp
* 29332 30052: contig of 721 bp in length
* 30053 30152: gap of 100 bp
* 30153 30859: contig of 707 bp in length
* 30860 30959: gap of 100 bp
* 30960 31677: contig of 718 bp in length
* 31678 31777: gap of 100 bp
* 31778 32493: contig of 716 bp in length
* 32494 32593: gap of 100 bp
* 32594 33298: contig of 705 bp in length
* 33299 33398: gap of 100 bp
* 33399 34116: contig of 718 bp in length
* 34117 34216: gap of 100 bp
* 34217 34943: contig of 727 bp in length
* 34944 35043: gap of 100 bp
* 35044 35765: contig of 722 bp in length
* 35766 35865: gap of 100 bp
* 35866 36588: contig of 723 bp in length
* 36589 37405: contig of 717 bp in length
* 37406 37505: gap of 100 bp
* 37506 38225: contig of 720 bp in length
* 38226 38325: gap of 100 bp
* 38326 39014: contig of 689 bp in length
* 39015 39114: gap of 100 bp
* 39115 39810: contig of 696 bp in length
* 39811 39910: gap of 100 bp
* 39911 40632: contig of 722 bp in length
* 40633 40732: gap of 100 bp
* 40733 41457: contig of 725 bp in length
* 41458 41557: gap of 100 bp
* 41558 42272: contig of 715 bp in length
* 42273 42372: gap of 100 bp
* 42373 43061: contig of 689 bp in length
* 43062 43161: gap of 100 bp
* 43162 43893: contig of 732 bp in length
* 43894 44690: gap of 100 bp
* 44690: contig of 697 bp in length


```

* 44691 44790: gap of 100 bp
* 44791 45485: contig of 695 bp in length
* 45486 45585: gap of 100 bp
* 45586 46302: contig of 717 bp in length
* 46303 46402: gap of 100 bp
* 46403 47105: contig of 703 bp in length
* 47106 47205: gap of 100 bp
* 47206 47922: contig of 717 bp in length
* 47923 48022: gap of 100 bp
* 48023 48751: contig of 729 bp in length
* 48752 48851: gap of 100 bp
* 48852 49549: contig of 698 bp in length
* 49550 49649: gap of 100 bp
* 49650 50364: contig of 715 bp in length
* 50365 50464: gap of 100 bp
* 50465 51182: contig of 718 bp in length
* 51183 52006: contig of 724 bp in length
* 52007 52106: gap of 100 bp
* 52107 52837: contig of 731 bp in length
* 52838 52937: gap of 100 bp
* 52938 53664: contig of 727 bp in length
* 53665 53764: gap of 100 bp
* 53765 54484: contig of 720 bp in length
* 54485 54584: gap of 100 bp
* 54585 55307: contig of 723 bp in length
* 55308 55407: gap of 100 bp
* 55408 56111: contig of 704 bp in length

```

```

Query Match      80.8%; Score 20.2; DB 14; Length 65091;
Best Local Similarity 88.0%; Pred. No. 9.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAACTAGCTGATCTT 25
Db 7562 AAAAAACATCTAGCTTATTTT 7586

```

```

RESULT 49
AP004509/c 84322 bp DNA linear PLN 22-JUL-2003
LOCUS Lotus corniculatus var. japonicus genomic DNA, chromosome 6,
DEFINITION clone:LT16K17, TM0037a, complete sequence.
ACCESSION AP004509
VERSION AP004509.1 GI:17736876
KEYWORDS HTG.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
1
REFERENCE 1
AUTHORS Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S.
TITLE Structural Analysis of a Lotus japonicus Genome. I. Sequence
Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb
Regions of the Genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84322)
AUTHORS Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935,
Fax: 81-438-52-3934)
FEATURES
Source
Location/Qualifiers
1. 84322
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"
/chromosome="6"

```

```

/clone="LT16K17"
/clone_lib="LT library"
/notes="TM0037a, a part of TAC clone: TM0037"
synonym: Lotus japonicus"

```

```

ORIGIN
Query Match      80.8%; Score 20.2; DB 15; Length 84322;
Best Local Similarity 88.0%; Pred. No. 9.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAACTAGCTGATCTT 25
Db 24440 AAAAAAACACAGACTTGATCTT 24416

```

```

RESULT 50
AC106873/c 99596 bp DNA linear PRI 31-JAN-2004
LOCUS Homo sapiens BAC clone RP11-374W7 from 7, complete sequence.
DEFINITION AC106873
ACCESSION AC106873
VERSION AC106873.3 GI:18698877
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 99596)

```

```

REFERENCE 1
AUTHORS Hillier, L.W., Fulton, R.S., Fulton, L.A., Graves, T.A., Pepin, K.H.,
Wagner-McPherson, C., Layman, D., Maas, J., Jaeger, S., Walker, R.,
Wylie, K., Sekhon, M., Becker, M.C., O'Laughlin, M.D., Schaller, M.E.,
Fowell, G.A., Delehaunty, K.D., Miner, T.L., Nash, W.E., Cordes, M.,
Du, H., Sun, H., Edwards, J., Bradshaw-Cordum, H., Ali, J., Andrews, S.,
Isack, A., Vanbrunt, A., Nguyen, C., Du, F., Lamar, B., Courtney, L.,
Kalicki, J., Ozersky, P., Bielicki, L., Scott, K., Holmes, A.,
Harkins, R., Harris, A., Strong, C.W., Hou, S., Tomlinson, C.,
Dauphin-Kohlberg, S., Kozlowicz-Reilly, A., Leonard, S., Rohlfing, T.,
Rock, S.M., Tin-Mollam, A.M., Abbott, A., Minx, P., Maupin, R.,
Stromwater, C., Latreille, P., Miller, N., Johnson, D., Murray, J.,
Woesener, J.P., Wendl, M.C., Yang, S.P., Schuler, B.R., Wallis, J.W.,
Spieeth, J., Bieri, T.A., Nelson, J.O., Berkowicz, N., Mohlmann, P.E.,
Cook, L.L., Hickenbotham, M.T., Eldred, J., Williams, D., Bedell, J.A.,
Wardis, E.R., Clifton, S.W., Chissole, S.J., Marra, M.A., Raymond, C.,
Haugen, E., Gillett, W., Zhou, Y., James, R., Phelps, K., Iadamoto, S.,
Bubb, K., Sims, E., Levy, R., Clendenning, J., Kau, R., Kent, W.J.,
Bork, P., Suyama, M., Bailey, J.A., Portnoy, M.E., Torrence, D.,
Chinaila, A.T., Gish, W.R., Eddy, S.R., McPherson, J.D., Olson, M.V.,
Eichler, E.E., Green, E.D., Waterston, R.H. and Wilson, R.K.
The DNA sequence of human chromosome 7
JOURNAL Nature 424 (6945), 157-164 (2003)
PUBMED 12853948

```

```

REFERENCE 2 (bases 1 to 99596)
AUTHORS Cotton, M. and Grewal, N.
TITLE The sequence of Homo sapiens BAC clone RP11-374W7
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 99596)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 99596)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 99596)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

REFERENCE 6 (bases 1 to 99596)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2004) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Feb 17, 2002 this sequence version replaced g1:18158390.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: sapiens@genome.wustl.edu
----- Summary Statistics
Center project name: H_NH0374M07

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, B., Tatenoe, M., Catanesse, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

Location/Qualifiers

FEATURES

source

1..99596
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-374M7"
/clone_lib="RPCI-11"
5..138
/rpt_family="Alu"
399..604
/rpt_family="L1"
605..919
/rpt_family="Alu"
920..1039
/rpt_family="L1"
1045..1325
/rpt_family="Alu"
1330..1444
/rpt_family="(GAAA)n"
1552..1572
/rpt_family="AT_rich"
1699..2061
/rpt_family="MER2_type"
2062..2284
repeat_region

repeat_region
2285..2351
/rpt_family="MER2_type"
2949..3225
/rpt_family="Alu"
3237..3413
/rpt_family="Alu"
3438..3548
/rpt_family="L1"
3549..3859
/rpt_family="Alu"
3860..3891
/rpt_family="L1"
3936..4009
/rpt_family="CR1"
5320..5362
/rpt_family="CR1"
5365..5446
/rpt_family="ACHobo"
5491..5521
/rpt_family="MIR"
5526..5571
/rpt_family="CR1"
complement(5771..5860)
/product="rRNA-5S_rRNA-related"
5771..5860
/rpt_family="5S"
6245..6555
/rpt_family="Alu"
6785..6812
/rpt_family="AT_rich"
7501..7800
/rpt_family="Alu"
7864..7884
/rpt_family="AT_rich"
8393..8707
/rpt_family="Alu"
9733..10028
/rpt_family="Alu"
10324..10521
/rpt_family="MIR"
11074..11279
/rpt_family="L2"
11336..11642
/rpt_family="Alu"
11690..12111
/rpt_family="CT-rich"
12115..12392
/rpt_family="Alu"
12924..13569
/rpt_family="ERV1"
14427..14642
/rpt_family="MIR"
14585..15717
/note="CGI_island (%GC=70.4, o/e=0.91, #CPGs=143)"
14788..15137
/gene="ST7"
14788..15137
/gene="ST7"
14814..14913
/rpt_family="GC_rich"
14987..>15137
/note="Homo sapiens suppression of tumorigenicity 7 (ST7), transcript variant a, mRNA; H_NH0374M07.1
This gene was based on g1(11761623 11761625)
continues as H_RG114A06.2.1"
/codon_start=1
/product="unknown"
/protein_id="AAS07529.1"
/db_xref="GI:41473937"

Query Match

80.8%; Score 20.2; DB 8; Length 99596;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 : Search time 203.2 Seconds
(without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-4
Perfect score: 25
Sequence: 1 aaaaaaaaaactatagcttgatcctt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database :

N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001as:*
6: geneseqn2002as:*
7: geneseqn2002as:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20.2	80.8	5228	6	ABL33654	Ab133654 Human imm
C 2	20.2	80.8	15649	6	ABL70543	Ab170543 Chemical
C 3	19.8	79.2	1666	13	ADX31754	Adx31754 Plant ful
C 4	19.4	77.6	209	4	AA182692	AA182692 Human pol
C 5	19.4	77.6	211	4	AA182693	AA182693 Human pol
C 6	19.4	77.6	276276	11	ACN44350	Acn44350 Human gen
C 7	19.2	76.8	467	3	AAA81560	Aaa81560 N. mening
C 8	19.2	76.8	541	13	AD554458	Ad554458 Bacteri
C 9	19.2	76.8	2133	13	AD551136	Ad551136 Bacteri
C 10	19.2	76.8	5423	6	ABN80000	Abn80000 Human che
C 11	19.2	76.8	6141	6	ABL09004	Ab109004 Drosophi
C 12	19.2	76.8	6247	6	ABK39923	Abk39923 Human che
C 13	19.2	76.8	7485	4	AB14740	Ab14740 Drosophi
C 14	19.2	76.8	10811	2	AAV59091	Aav59091 Potato po
C 15	19.2	76.8	13606	4	AA545458	Aa545458 Chemical
C 16	19.2	76.8	13606	4	AA545458	Aa545458 Chemical
C 17	19.2	76.8	13606	6	ABL33811	Ab133811 Human imm
C 18	19.2	76.8	13606	6	ABK28314	Abk28314 DNA trans
C 19	19.2	76.8	14032	6	ABL33452	Ab133452 Human imm

C 20	19.2	76.8	139904	6	ABK83562	Abk83562 Human CDN
C 21	19.2	76.0	1114	2	AA296225	Aa296225 S. pneumo
C 22	19.2	76.0	1117	2	AAT98574	Aat98574 DNA encod
C 23	18.8	75.2	460	4	AAK90319	Aak90319 Human dig
C 24	18.8	75.2	460	4	AA157693	Aa157693 Human col
C 25	18.8	75.2	460	6	AB599870	Ab599870 Genomic D
C 26	18.8	75.2	460	10	ADB93023	Adb93023 Human col
C 27	18.8	75.2	469	4	AAK88382	Aak88382 Human dig
C 28	18.8	75.2	469	4	AA157594	Aa157594 Human col
C 29	18.8	75.2	469	6	AB599771	Ab599771 CDNA enco
C 30	18.8	75.2	469	10	ADB92851	Adb92851 Human col
C 31	18.8	75.2	573	13	ACN61173	Acn61173 Cotton gy
C 32	18.8	75.2	6749	4	AA546525	Aa546525 Tumour su
C 33	18.8	75.2	7348	4	AA546336	Aa546336 Tumour su
C 34	18.8	75.2	7703	6	ABL33435	Ab133435 Human imm
C 35	18.8	75.2	9741	6	ABL33323	Ab133323 Human imm
C 36	18.8	75.2	12987	2	AAT91324	Aat91324 Arabidops
C 37	18.8	75.2	47038	13	ACN37235	Acn37235 Human per
C 38	18.8	75.2	209083	13	ABD32854	Abd32854 Human can
C 39	18.6	74.4	312	5	ABV06607	Abv06607 Human pro
C 40	18.6	74.4	391	5	ABV36556	Abv36556 Human pro
C 41	18.6	74.4	424	6	ABV04081	Abv04081 Human pro
C 42	18.6	74.4	433	6	ABN21922	Abn21922 Human ORF
C 43	18.6	74.4	438	2	AAV90433	Aav90433 EST clone
C 44	18.6	74.4	664	13	ADK53399	Adk53399 Plant ful
C 45	18.6	74.4	703	6	ABQ35845	Abq35845 Oligonuc1
C 46	18.6	74.4	703	6	ABQ35844	Abq35844 Oligonuc1
C 47	18.6	74.4	774	4	ABL15043	Ab115043 Drosophi1
C 48	18.6	74.4	1157	3	AAK33525	Aak33525 Arabidops
C 49	18.6	74.4	1160	6	ABL34328	Ab134328 Human imm
C 50	18.6	74.4	1210	6	ABV83585	Abv83585 Human bre
C 51	18.6	74.4	1257	11	AD236132	Ad236132 Chinese c
C 52	18.6	74.4	1257	11	AD236131	Ad236131 Mustard t
C 53	18.6	74.4	2269	4	ADL62004	Adl62004 Human ova
C 54	18.6	74.4	2954	5	ABL15042	Ab115042 Drosophi1
C 55	18.6	74.4	3011	12	ADQ36007	Adq36007 Novel mou
C 56	18.6	74.4	3360	14	ADY30450	Ady30450 Human spl
C 57	18.6	74.4	3990	14	ADY30449	Ady30449 Human spl
C 58	18.6	74.4	4254	4	ABL04842	Ab104842 Drosophi1
C 59	18.6	74.4	4960	14	ADK06233	Adk06233 Cyclin-de
C 60	18.6	74.4	5003	3	AAK77125	Aak77125 Human ORF
C 61	18.6	74.4	5008	10	ADQ32472	Adq32472 Human nov
C 62	18.6	74.4	5040	12	ADQ23943	Adq23943 Human sof
C 63	18.6	74.4	5581	4	ABL06750	Ab106750 Drosophi1
C 64	18.6	74.4	5730	4	AB120024	Ab120024 Drosophi1
C 65	18.6	74.4	5962	6	ABL33469	Ab133469 Human imm
C 66	18.6	74.4	6050	6	ABL34011	Ab134011 Human imm
C 67	18.6	74.4	6051	12	ADQ21014	Adq21014 Human sof
C 68	18.6	74.4	6134	4	AA546564	Aa546564 Tumour su
C 69	18.6	74.4	6287	4	ABL18016	Ab118016 Drosophi1
C 70	18.6	74.4	6345	6	ABL33771	Ab133771 Human imm
C 71	18.6	74.4	6478	10	ADB54284	Adb54284 Pretreat
C 72	18.6	74.4	6478	13	AD589582	Ad589582 Oligonuc1
C 73	18.6	74.4	6478	13	AD589308	Ad589308 Oligonuc1
C 74	18.6	74.4	6478	6	ABK28270	Abk28270 DNA trans
C 75	18.6	74.4	6478	6	ABN80201	Abn80201 Human can
C 76	18.6	74.4	6746	4	ABL03756	Ab103756 Drosophi1
C 77	18.6	74.4	6746	4	AA546688	Aa546688 Tumour su
C 78	18.6	74.4	7303	6	ABK13103	Abk13103 Signal tr
C 79	18.6	74.4	7461	6	ABL33784	Ab133784 Human imm
C 80	18.6	74.4	7631	6	ABL32860	Ab132860 Human imm
C 81	18.6	74.4	7631	6	ABL11972	Ab111972 Drosophi1
C 82	18.6	74.4	8333	4	AA545406	Aa545406 Chemical
C 83	18.6	74.4	8333	6	ABL33502	Ab133502 Human imm
C 84	18.6	74.4	8333	6	ABK28255	Abk28255 DNA trans
C 85	18.6	74.4	9182	4	ABL04532	Ab104532 Drosophi1
C 86	18.6	74.4	9182	6	ABL32377	Ab132377 Human imm
C 87	18.6	74.4	9483	6	ABK70516	Abk70516 Chemical
C 88	18.6	74.4	9483	6	AA561092	Aa561092 Human gen
C 89	18.6	74.4	9483	6	AA561092	Aa561092 Human gen
C 90	18.6	74.4	9483	6	AA561092	Aa561092 Human gen
C 91	18.6	74.4	9483	6	AA561092	Aa561092 Human gen
C 92	18.6	74.4	9483	6	AA561092	Aa561092 Human gen

C 93	18.6	74.4	9725	6	ABLJ3292	Ab13292 Human imm
C 94	18.6	74.4	9725	6	ABN0180	Abn0180 Human che
C 95	18.6	74.4	10006	6	ABLJ2037	Ab12037 Human imm
C 96	18.6	74.4	11490	10	ADG32973	Adg32973 Human DNA
C 97	18.6	74.4	11490	13	ADR25619	Adr25619 Breast ca
C 98	18.6	74.4	11490	14	ADY61850	Ady61850 Human gen
C 99	18.6	74.4	11836	4	AA545394	AA545394 Chemical
C 100	18.6	74.4	11836	6	ABK28239	Abk28239 DNA trans
C 101	18.6	74.4	12046	6	ABLJ3632	Ab13632 Human imm
C 102	18.6	74.4	12792	5	AAH20176	Aah20176 Human mut
C 103	18.6	74.4	12793	5	AAH20182	Aah20182 Human mut
C 104	18.6	74.4	12793	5	AAH20178	Aah20178 Human mut
C 105	18.6	74.4	12793	5	AAH20174	Aah20174 Human spa
C 106	18.6	74.4	12793	5	AAH20179	Aah20179 Human mut
C 107	18.6	74.4	12793	12	ADQ19317	Adq19317 Human sof
C 108	18.6	74.4	12793	14	ADX05804	Adx05804 Cyclin-de
C 109	18.6	74.4	12892	12	ADQ23573	Adq23573 Human sof
C 110	18.6	74.4	13085	4	AB103680	Ab103680 Drosophill
C 111	18.6	74.4	14000	3	AAA09020	Aaa09020 A. thailia
C 112	18.6	74.4	17216	5	ABA137200	Ab137200 Human mus
C 113	18.6	74.4	17216	5	ABA18237	Ab18237 Human ner
C 114	18.6	74.4	17216	8	ABX60188	Abx60188 cDNA enco
C 115	18.6	74.4	17216	12	ADJ30938	Adj30938 Human mus
C 116	18.6	74.4	17217	4	AA137201	Aa137201 Human mus
C 117	18.6	74.4	17217	5	ABA18239	Ab18239 Human ner
C 118	18.6	74.4	17217	8	ABX60189	Abx60189 cDNA enco
C 119	18.6	74.4	17217	12	ADJ30939	Adj30939 Human mus
C 120	18.6	74.4	17703	6	ABK39952	Abk39952 Human che
C 121	18.6	74.4	17869	6	ABK39920	Abk39920 Human che
C 122	18.6	74.4	17869	6	ABLJ32104	Ab132104 Human imm
C 123	18.6	74.4	18283	6	ABL70502	Ab170502 Chemical
C 124	18.6	74.4	18283	6	AA561363	Aa561363 Human gen
C 125	18.6	74.4	18598	6	ABLJ2387	Ab12387 Human imm
C 126	18.6	74.4	21737	4	ABL06944	Ab106944 Drosophill
C 127	18.6	74.4	22636	4	AAK81855	Aak81855 Human imm
C 128	18.6	74.4	23253	4	AB106900	Ab106900 Drosophill
C 129	18.6	74.4	23254	4	AA105829	Aa105829 Human rep
C 130	18.6	74.4	23544	4	ABL89393	Ab189393 Human tes
C 131	18.6	74.4	26230	11	ACN44678	Acn44678 Human gen
C 132	18.6	74.4	26371	11	ACN44814	Acn44814 Human gen
C 133	18.6	74.4	32206	4	AAK89374	Aak89374 Human dig
C 134	18.6	74.4	32433	9	ADA02630	Ada02630 Human FLT
C 135	18.6	74.4	32433	10	ADB72368	Adb72368 Human FLT
C 136	18.6	74.4	32433	10	ADB5878	Adb5878 Human FLT
C 137	18.6	74.4	39401	14	ADJ13218	Adj13218 Murine ca
C 138	18.6	74.4	41991	13	ABD32684	Abd32684 Human can
C 139	18.6	74.4	50927	13	ABD32585	Abd32585 Human can
C 140	18.6	74.4	61020	4	AA546788	Aa546788 Tumour su
C 141	18.6	74.4	62154	12	ADQ97611	Adq97611 Human can
C 142	18.6	74.4	101433	14	AEA6169_3	Continuation (4 of
C 143	18.6	74.4	110000	8	ABX16390_1	Continuation (2 of
C 144	18.6	74.4	117730	14	ADJ12550	Adj12550 Human can
C 145	18.6	74.4	169739	6	ABO88186	Abc88186 Human oar
C 146	18.6	74.4	177566	14	ABE71426	Aeb71426 Human oar
C 147	18.6	74.4	177563	9	ACD28257	Acg28257 Mouse sol
C 148	18.6	74.4	185371	6	AB110718	Abt10718 Human bre
C 149	18.6	74.4	235070	11	ACN45174	Acn45174 Human gen
C 150	18.6	74.4	264965	12	ADN16203	Adn16203 Human sul

ALIGNMENTS

RESULT 1
ID ABLJ3654/c
ABLJ3654 standard; DNA; 5228 BP.

XX ABLJ3654;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1627.

XX

KW	Human; immune system disease; cytosine methylation; antiasthmatic;
KW	antiarteriosclerotic; antiataemic; cytosatic; nocrotopic;
KW	antiprotective; anti-HIV; anticonvulsant; ophthalmological;
KW	antirheumatic; antirachitic; antidiabetic; antipsoriatic;
KW	antiflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW	ds.
OS	Homo sapiens.
XX	
XX	MO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP007537.
XX	
PR	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
XX	
PA	(EPIC-) EPIGENOMICS AG.
XX	
PI	Olek A, Piegenbrock C, Berlin K;
XX	
DR	WPI, 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful for
PT	diagnosis and treatment of diseases associated with abnormal cytosine
PT	methylation.
XX	
PS	Claim 1; SEQ ID NO 1627; 32PP + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	gene which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention
XX	
SO	Sequence 5228 BP; 1745 A; 37 C; 1156 G; 2290 T; 0 U; 0 Other;
XX	
QY	Query Match 80.8%; Score 20.2; DB 6; Length 5228;
XX	
DB	Best Local Similarity 88.0%; Pred. No. 3.1e+02;
XX	
XX	Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX	
QY	1 AAAAAAAAACTATACCTTGATCTT 25
XX	
DB	336 AAAAAAAAACTATACCTTTTCTT 312
XX	
XX	
RESULT 2	
ID	ABL70543/c
AB	ABL70543 standard; DNA; 15649 BP.
XX	
AC	ABL70543;
XX	
DT	01-JUL-2002 (first entry)
XX	
XX	
DE	Chemically treated cell signalling DNA sequence#217.
XX	
KW	Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW	tumour; cytosatic; ds.
XX	
OS	Unidentified.
XX	
XX	WO200202807-A2.
XX	
PN	10-JAN-2002.
XX	
PD	
XX	
PF	29-JUN-2001; 2001WO-EP007471.
XX	

PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K,
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX
PS Claim 1; SEQ ID NO 433; 24pp + Sequence Listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signaling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signaling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signaling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABU70111-ABU70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signaling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 15649 BP; 4938 A; 109 C; 3382 G; 7220 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 6; Length 15649;
Best Local Similarity 88.0%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTTGATCTT 25
Db 5501 AAAAAAAAACTATAGCTTGATCTT 5477
RESULT 3
ADX31754
ID ADX31754 standard; cDNA; 1666 BP.
XX
AC ADX31754;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polynucleotide seqid 14574.
XX
KM plant protectant; plant growth regulant; gene therapy; plant;
KM recombinant DNA construct; physical array; plant breeding marker;
KM cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KM extreme osmotic condition; pathogen tolerance; pest tolerance;
KM growth rate; cell cycle pathway; disease resistance;
KM galactomannan production; lignin production; plant growth regulator;
KM yield; plant growth; plant development; seed oil; protein yield;
KM protein content; gene; ss.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
XX 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX

PA (LITU/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABAS/) TABASKA J E.
PA (CAOV/) CAO Y.
XX
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX
DR WPI; 2004-180133/17.
XX
PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 14574; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?docid:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 1666 BP; 532 A; 259 C; 429 G; 446 T; 0 U; 0 Other;
Query Match 79.2%; Score 19.8; DB 13; Length 1666;
Best Local Similarity 91.3%; Pred. No. 4.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 AAAAAAAAACTATAGCTTGATCTT 25
Db 1643 AAAAAAAAACTATAGCTTGATCTT 1665
RESULT 4
AA182692/c
ID AA182692 standard; cDNA; 209 BP.
XX
AC AA182692;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2752.
XX
KM Human; cytokines; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX

```

XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
DR P-PSDB; AAO02761.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 1; SEQ ID NO 2752; 1399pp + Sequence listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activating/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 209 BP; 42 A; 58 C; 42 G; 67 T; 0 U; 0 Other;
SQ
XX
XX Query Match 77.6%; Score 19.4; DB 4; Length 209;
XX Best Local Similarity 95.2%; Pred. No. 5.7e+02;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAACTATAGCTTGA 21
QY |||||
XX 74 AAAAAAAAACTATAGCTTGA 54
DB
XX
XX RESULT 5
XX AA182693/c
XX ID AA182693 standard; cDNA; 211 BP.
XX
XX AA182693;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polynucleotide SEQ ID NO 2753.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200164835-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001MO-US004927.
PF
XX
XX 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI; 2001-514838/56.
DR P-PSDB; AAO02762.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.

```

```

XX Claim 1; SEQ ID NO 2753; 1399pp + Sequence listing; English.
PS
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activating/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 211 BP; 42 A; 56 C; 47 G; 66 T; 0 U; 0 Other;
SQ
XX
XX Query Match 77.6%; Score 19.4; DB 4; Length 211;
XX Best Local Similarity 95.2%; Pred. No. 5.7e+02;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAACTATAGCTTGA 21
QY |||||
XX 74 AAAAAAAAACTATAGCTTGA 54
DB
XX
XX RESULT 6
XX ACN44350
XX ID ACN44350 standard; DNA; 276276 BP.
XX
XX ACN44350;
AC
XX
XX 18-NOV-2004 (first entry)
DT
XX
XX Human genomic sequence hCG17121.
DE
XX
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX MO2003073826-A2.
PN
XX
XX 12-SEP-2003.
PD
XX
XX 28-FEB-2003; 2003WO-US006235.
PF
XX
XX 01-MAR-2002; 2002US-00087192.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
XX Morris DW;
PI
XX
XX WPI; 2003-328604/31.
DR
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
XX Claim 1; SEQ ID NO 754; 0pp; English.
PS
XX
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a bioclip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding

```


CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 276276 BP, 68379 A; 69211 C; 66764 G; 71922 T; 0 U; 0 Other;
Query Match 77.6%; Score 19.4; DB 11; Length 276276;
Best Local Similarity 95.2%; Pred. No. 6.7e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTTGA 21
Db 17567 AAAAAAAAACTATAGCTAGA 17587
RESULT 7
AAAB1560/c
ID AAAB1560 standard; DNA; 467 BP.
XX
AC AAAB1560;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_107 SEQ ID NO:107.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US023573.
XX
PR 09-OCT-1998; 98US-0103794P.
PR 30-APR-1999; 99US-0132068P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarcelli M, Scarlato V;
PI Rappoli R, Pizza M;
XX
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be used
XX in the diagnosis and treatment of N. meningitidis infection and other
XX Neisserial infections, for example, N.gonorrhoea.
XX
XX Claim 7; Page 1496; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic proteins
XX from Neisseria genomic sequences. AAAB1453 to AAAB2414 represent
XX specifically claimed Neisseria meningitidis genomic DNA sequences;
XX AAAB1260 to AAAB1303 and AAB25620 to AAB25663 represent Neisseria DNA
XX sequences and their corresponding proteins; AAAB1254 to AAAB1259 and
XX AAAB1304 to AAAB1321 represent PCR primers used in the isolation of
XX Neisseria meningitidis DNA sequences; and AAAB1322 to AAAB1452 represent
XX Neisseria meningitidis MenB polynucleotide ORF sequences, which are all
XX used in the exemplification of the present invention. The nucleic acid
XX sequences, protein sequences, and antibodies against them, can be used in
XX the manufacture of a composition. The composition can be used as a
XX medicament (or in the manufacture of a medicament) for treating,
XX preventing or diagnosing infection due to Neisserial bacteria. For
XX example, some of the identified proteins could be components of vaccines
XX against Meningococcus B; against all serotypes; and/or against all
XX pathogenic Neisseriae. Identification of sequences from the Bacterium
XX will also facilitate production of biological probes, particularly
XX organism-specific probes. Attempts to make efficacious Meningococcus B
XX vaccines have failed mainly due to antigen tolerance. Multivalent
XX vaccines have also been tried but none have successfully overcome

CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
CC variable regions
XX
SQ Sequence 467 BP, 163 A; 67 C; 55 G; 182 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 3; Length 467;
Best Local Similarity 87.5%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTGATCT 24
Db 233 AAAAAAAAACTATAGCATTAGCT 210
RESULT 8
ADS54458/c
ID ADS54458 standard; cDNA; 541 BP.
XX
AC ADS54458;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #6445.
XX
KM Recombinant DNA construct; transformed plant; improved plant property;
KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KM pathogen tolerance; pest tolerance; plant disease resistance;
KM cell cycle pathway modification; plant growth regulator;
KM homologous recombination; seed oil yield; protein yield; carbohydrate;
KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KM bacterial polynucleotide; gene; ss.
XX
XX Bacteria.
OS
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 30132; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX CC promoter functional in a plant cell, where the promoter is positioned to
XX CC provide for expression of a polynucleotide encoding a polypeptide from a
XX CC microbial source. The invention also relates to a transformed plant
XX CC comprising the recombinant DNA construct and a method of producing a
XX CC transforming plant having an improved property. The plant is a crop plant
XX CC such as maize or soybean. The method of producing a transformed plant
XX CC having an improved property comprises transforming a plant with the
XX CC recombinant DNA construct and growing the transformed plant, where the
XX CC polynucleotide or polypeptide is useful for improving plant properties.
XX CC The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,

CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or by
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production, This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 541 BP; 229 A; 75 C; 128 G; 109 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 13; Length 541;
Best Local Similarity 87.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTTGATCT 24
DB 294 AAAAAAAAAATTATAGATTGATCT 271
RESULT 9
ID ADS51136 standard; cDNA; 2123 BP.
XX
AC ADS51136;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #5879.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX pathogen tolerance; pest tolerance; plant disease resistance;
XX cell cycle pathway modification; plant growth regulator;
XX homologous recombination; seed oil yield; protein yield; carbohydrate;
XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR MPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 29566; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a

CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production, This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2123 BP; 782 A; 272 C; 308 G; 761 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 13; Length 2123;
Best Local Similarity 87.5%; Pred. No. 7.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTTGATCT 24
DB 353 AAAAAAAAACTATTTCTTGCTCT 376
RESULT 10
ID ABN80000/c
XX
AC ABN80000;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 17.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
XX heart disease; epilepsy; histone deacetylation; muscular dystrophy;
XX dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
XX antidiabetic; cytostatic; anticonvulsant; de.
XX
OS Homo sapiens.
XX
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR MPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
XX associated with development genes such as diabetes, comprises a sequence
XX of a segment of chemically pretreated DNA of genes associated with
XX development.
XX
PS Claim 1; SEQ ID NO 17; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)

CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACPN, ADPN, or APD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Curranio syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (ii) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (iii) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 5432 BP; 1381 A; 116 C; 1240 G; 2695 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 6; Length 5432;
Best Local Similarity 87.5%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 2743 AAAAAAAAACTATAGCTTGATCT 2720

RESULT 11
ID ABL09004
AC ABL09004 standard; cDNA; 6141 BP.
XX
XX ABL09004;
XX
XX 26-MAR-2002 (first entry)
XX
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 21494.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX W0200171042-A2.
XX
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX
XX Venter JC, Adams M, Li FWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB64901.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX
XX Claim 1; SEQ ID NO 21494; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-

CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6141 BP; 1715 A; 1266 C; 1288 G; 1872 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 4; Length 6141;
Best Local Similarity 87.5%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 AAAAAAAAACTATAGCTTGATCTT 25
Db 4636 AAAAAAAAACTATAGCTTGATTT 4659

RESULT 12
ID ABR39923/c
AC ABR39923 standard; DNA; 6247 BP.
XX
XX
XX 21-MAY-2002 (first entry)
XX
XX
XX Human chemically pretreated gene sequence #2 strand 2.
DE
XX
XX Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
XX
XX cytosine; ALDH6; CYP1A; CYP1B1; CYP3A3; DYPD; EPHX2; OCLN; TXNRD1;
XX
XX UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
XX Homo sapiens.
OS
XX
XX W0200202806-A2.
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-EP007470.
XX
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPig-) EPIGENOMICS AG.
XX
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-154757/20.
XX
XX
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
XX PT useful for detecting cytosine methylation state of genes associated with
XX PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
XX
XX Claim 1; SEQ ID NO 4; 24pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence at least 18
XX bases in length of a segment of the chemically pretreated DNA of genes
XX CC associated with pharmacogenomics according to one of the sequences of the
XX CC genes ALDH6 (NM 000693), CYP1A (NM 000781), CYP1B1 (NM 000497), CYP3A3
XX CC (NM 000776 and NM 017460), DYPD (NM 000110), EPHX2 (NM 001979), OCLN
XX CC (NM 002538), TXNRD1 (NM 003330), UGT8 (NM 003360), MRP (NM 004996,
XX CC NM 019900, NM 019901, NM 019902, NM 019862, NM 019898, NM 019899) and
XX CC their complementary sequences, or a sequence (51) chosen from 87
XX CC sequences and their complements. The chemical pretreatment is bisulphite
XX CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
XX CC Also included are an oligomer (ii) in particular an oligonucleotide or a
XX CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
XX CC base sequence having a length of 9 nucleotides which hybridises to or is
XX CC identical to a chemically pretreated DNA of genes associated with
XX CC pharmacogenomics and their complements, arranged in an array for
XX CC analysing diseases associated with the methylation state (CpG) and/or
XX CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
XX CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
XX CC and their complements is useful for diagnosis and therapy of solid
XX CC tumours and cancer. The present sequence represents one the 87 DNA
XX CC sequences or its complement. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6247 BP: 1592 A; 143 C; 1378 G; 3134 T; 0 U; 0 Other;

Query Match	76.8%	Score 19.2;	DB 6;	Length 6247;
Best Local Similarity	87.5%	Pred. No. 7.4e+02;		
Matches 21; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

QY 1 AAAAAAAAACTATAGCTTGATCT 24
| | | | | | | | | | | |
Db 2715 AAAAAAAATCTATACCTTTATCT 2692

RESULT 13
ABL14740
ID ABL14740 standard; cDNA; 7495 BP.

Query Match	75.8%	Score 19.2	DB 4	Length 7495
Best Local Similarity	87.5%	Pred. No. 7.4e+02		
Matches 21; Conservative	0	Mismatches 3	Indels 0	Gaps 0

QY	1	AAAAAAAAAACTATAGCTTGATCT	24
Db	3902	ATAAAGAACTATAGCTTGA	3925

RESULT 14
AAV59091
ID AAV59091 standard; DNA; 10811 BP.

Query Match	76.8%	Score 19.2;	DB 2;	Length 10811;
Best Local Similarity	87.5%;	Pred. No. 7.5e+02;		
Matches 21; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

```
QY      1 AAAAAAAAACTATAGCTGATCT 24  
        |||||  
Db     10251 AAAAAAAAACTATGGATTAACTC 10274
```

RESULT 15
AAS45458/c
ID AAS45458 standard; DNA; 13606 BP.

KM immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KM PCR primer.
 XX
 OS Homo sapiens.
 PN WO200168911-A2.
 XX
 PD 20-SEP-2001.
 PF
 XX 15-MAR-2001; 2001WO-EP002945.
 XX
 PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602751/68.
 XX
 PT Designing primers and probes for analyzing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle.
 XX
 PS Claim 1; SEQ ID NO 163; 28pp; English.
 XX
 CC Sequences AAS45396-AAS45530 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers
 CC
 SQ Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
 Query Match 76.8%; Score 19.2; DB 4; Length 13606;
 Best Local Similarity 87.5%; Pred. No. 7.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTATAGCTTGATCT 24
 Db 5439 AAAAAAAAACTAAACTTAATCT 5416
 RESULT 16
 AAS4562/c
 ID AAS4562 standard; DNA; 13606 BP.
 XX
 AC AAS4562;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #284.
 XX
 KM Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
 KM tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KM cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.

XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP002955.
 XX
 PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumor suppressor
 PT genes and oncogenes, useful in designing primers and probes for analyzing
 PT diseases associated with cytosine methylation state e.g. cancer.
 XX
 PS Claim 1; SEQ ID NO 284; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and oncogenes
 CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
 CC 500 are missing from the sequence listing) sequences (ss) and sequences
 CC complementary to (ss). The nucleic acid may be a peptide nucleic acid-
 CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
 CC probes for detecting the cytosine methylation state and/or single
 CC nucleotide polymorphisms and also to be used in an array for analysing
 CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
 CC probes can also be used in a method for ascertaining genetic and/or
 CC epigenetic parameters for the diagnosis and/or therapy of existing
 CC diseases or the predisposition to specific diseases, by analysing
 CC cytosine methylations. The parameters may be compared to another set of
 CC genetic and/or epigenetic parameters, the differences serving as basis
 CC for diagnosis and/or prognosis events which are disadvantageous to
 CC patients. The present sequence is one of the 533 genomic sequences
 CC derived from tumour suppressor genes and oncogenes. Sequences with even
 CC numbered Seq ID numbers are the complementary sequence of the
 CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
 CC 535, except for those whose partner sequence is missing). Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
 Query Match 76.8%; Score 19.2; DB 4; Length 13606;
 Best Local Similarity 87.5%; Pred. No. 7.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTATAGCTTGATCT 24
 Db 5439 AAAAAAAAACTAAACTTAATCT 5416
 RESULT 17
 ABL33811/c
 ID ABL33811 standard; DNA; 13606 BP.
 XX
 AC ABL33811;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1784.
 XX
 KM Human; immune system disease; cytosine methylation; antiasthmatic;
 KM antiarteriosclerotic; anti-anemic; cytostatic; neurotropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antineumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.
 XX Homo sapiens.
 OS
 XX
 XX MO200200928-A2.
 PN
 XX
 XX 03-JAN-2002.
 PD
 XX
 XX 02-JUL-2001; 2001WO-EP007537.
 PF
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 DR
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PS
 XX
 PS Claim 1, SEQ ID NO 1784, 32pp + Sequence Listing; German.
 CC
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/neuroinfective bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 XX
 SQ Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
 Query Match 76.8%; Score 19.2; DB 6; Length 13606;
 Best Local Similarity 87.5%; Pred. No. 7.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTATAGCTTGATCT 24
 DB 5439 AAAAAAAAACTATAAATTATCT 5416
 RESULT 18
 ID ABR28314/c
 ID ABR28314 standard; DNA; 13606 BP.
 XX
 AC ABR28314;
 XX
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE DNA transcription associated complementary genomic DNA #94.
 XX
 XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Maardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angioedema; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.
 XX
 XX Unidentified.
 OS
 XX
 XX MO200192565-A2.
 PN
 XX

PD 06-DEC-2001.
 XX
 XX 06-APR-2001; 2001WO-EP003973.
 PF
 XX
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-090046/12.
 DR
 XX
 XX New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological disorders,
 PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
 PT cancer.
 PS
 XX
 PS Claim 1, SEQ ID NO 188, 32pp; English.
 CC
 XX
 CC The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Maardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angioedema, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABR28127-ABR28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification but
 CC was obtained in electronic format directly from the European Patent
 CC Office
 CC
 SQ Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;
 Query Match 76.8%; Score 19.2; DB 6; Length 13606;
 Best Local Similarity 87.5%; Pred. No. 7.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTATAGCTTGATCT 24
 DB 5439 AAAAAAAAACTATAAATTATCT 5416
 RESULT 19
 ID ABL33452/c
 ID ABL33452 standard; DNA; 14032 BP.
 XX
 AC ABL33452;
 XX
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1425.
 XX
 XX Human immune system associated gene
 KW Human; immune system disease; cytosine methylation; antiaethmatic;
 KW antiarteriosclerotic; antihaemic; cyostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;

KM de.
 OS Homo sapiens.
 XX WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP007337.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 XX 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPiG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 DR
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PS
 XX Claim 1; SEQ ID NO 1425; 32pp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 SQ Sequence 14032 BP; 2890 A; 291 C; 4148 G; 6703 T; 0 U; 0 Other;
 QY
 DB 2 AAAAAAAAACTATGCTTGATCTT 25
 10064 AAAAAAAAACTATGCTTGATTTT 10041
 RESULT 20
 ABR83562/c
 ID ABR83562 standard; cDNA; 139904 BP.
 XX
 AC ABR83562;
 XX
 DT 14-AUG-2002 (first entry)
 DT
 DE Human cDNA differentially expressed in granulocytic cells #133.
 XX
 XX Human cDNA differentially expressed in granulocytic cells #133.
 KM Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KM viral infection; parasitic infection; protozoal infection;
 KM fungal infection; sterile inflammation; disease; psoriasis;
 KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KM adult respiratory distress syndrome; inflammatory bowel disease;
 KM Crohn's disease; ulcerative colitis; periodontal disease;
 KM granulocyte activation; chronic inflammation; allergy.
 KM
 XX Homo sapiens.
 OS
 AC
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US030821.
 XX
 PR 03-OCT-2000; 2000US-0237189P.

XX
 PA (GENE-) GENE LOGIC INC.
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX WPI; 2002-435328/46.
 DR
 XX
 PT Detecting granulocyte activation by detecting differential expression of
 PT genes associated with granulocyte activation, which serves as diagnostic
 PT markers that is useful for monitoring disease states and drug toxicity.
 PS
 XX Claim 1; SEQ ID NO 133; 114pp; English.
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing the
 CC expression level to an expression level in an unactivated GC, where
 CC differential expression of Gs is indicative of GCA. Also included are
 CC modulating (M2) GA by contacting GC with an agent that alters the
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent
 CC capable of modulating GCA or an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease using the gene expression
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease, by detecting the level of
 CC expression in a sample of the tissue of gene(s) from Gs, where the level
 CC of expression of the gene is indicative of inflammation; (4) treating
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease, by contacting a tissue having inflammation with an
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
 CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
 CC for screening an agent capable of modulating GCA preferably in an
 CC inflammation in a tissue; M4 is useful for detecting an inflammation
 CC (especially chronic) in a tissue, an allergic response in a subject,
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
 CC disease, ulcerative colitis, periodontal disease; also bacterial
 CC infection, viral infection, parasitic infection, protozoal infection,
 CC fungal infection and M5 is useful for treating one of the above
 CC conditions. The present sequence represents a gene differentially
 CC expressed in granulocytes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 139904 BP; 39268 A; 29759 C; 30173 G; 40704 T; 0 U; 0 Other;
 QY
 DB 1 AAAAAAAAACTATGCTTGATCTT 24
 36817 AAAAAAAAACTTACCTTGCTGT 36794
 RESULT 21
 AA296225
 ID AA296225 standard; DNA; 1114 BP.
 XX
 AC AA296225;
 XX
 DT 10-APR-2000 (first entry)
 DT
 DE S. pneumoniae derived DNA from ORF #53.
 XX
 PF Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KM bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
 XX

```

OS Streptococcus pneumoniae.
XX
XX WO9806734-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US014436.
XX
XX 16-AUG-1996; 96US-0024022P.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX
XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
XX Stodola RK;
XX
XX MPI; 1998-159452/14.
XX
XX P-PSDB; AAY85847.
XX
XX Streptococcus pneumoniae proteins and related DNA - useful for screening
XX compounds for antibacterial activity.
XX
XX Claim 4; Page 86-87; 640pp; English.
XX
XX This invention describes novel isolated Streptococcus pneumoniae (see
XX CC polymuclonides (see AAY86173-286494) and their encoded proteins (see
XX CC AAY8792-Y86182). The DNA, vectors and host cells described in the method
XX CC of the invention are useful for the recombinant expression of the
XX CC polypeptides. The polypeptides are useful for treatment or prevention of
XX CC disease, or diagnosis of disease related to expression or activity of
XX CC such a polypeptide. They can also be used to screen for compounds which
XX CC interact with and inhibit or activate such a polypeptide. The
XX CC polypeptides (or DNA encoding them, via gene therapy) are also useful for
XX CC inducing an immunological response in a mammal. The antagonists are
XX CC useful to inhibit such bacterial polypeptides. The polypeptides are
XX CC particularly useful to identify antimicrobial compounds and antibiotics.
XX CC They are also useful to determine their role in pathogenesis of
XX CC infection, dysfunction and disease
XX
XX Sequence 1114 BP; 308 A; 221 C; 229 G; 356 T; 0 U; 0 Other;
SQ
Query Match 76.0%; Score 19; DB 2; Length 1114;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AAAAAAAAACTATAGCTTG 20
Db 1011 AAAAAAAAACTATAGCTTG 1029

```

```

PD 20-NOV-1997.
XX
XX 14-MAY-1997; 97WO-US007950.
XX
XX 14-MAY-1996; 96US-0017670P.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO, Stodola RK;
XX
XX MPI; 1998-008793/01.
XX
XX P-PSDB; AAN38507.
XX
XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
XX PT diagnosing anti-microbial agents for treatment of bacterial infections.
XX
XX Claim 4; Page 111; 483pp; English.
XX
XX This sequence encodes a Streptococcus pneumoniae protein that (based on
XX CC homology with a Lactobacillus helveticus protein) is an ISL2 protein, and
XX CC represents a DNA sequence of the invention. The DNA sequences were
XX CC isolated from Streptococcus pneumoniae strain 0100993 (NCIMB 40794). The
XX CC Streptococcus pneumoniae proteins of the invention can be used to
XX CC identify compounds which interact with and inhibit or activate the
XX CC activity of the proteins. Antagonists can be used to treat diseases
XX CC caused by S. pneumoniae proteins, through genetic immunisation. They can
XX CC also be used to induce an immunological response in a mammal by
XX CC inoculation with the S. pneumoniae proteins or delivery of the encoding
XX CC nucleic acids in a vector adequate to produce antibody and/or T cell
XX CC immune responses to protect the animal from disease. The proteins can
XX CC also be used to identify antimicrobial compounds which are capable of
XX CC inhibiting their bioactivity. In particular the proteins of the invention
XX CC can be used to prevent adhesion of bacteria to mammalian extracellular
XX CC matrix proteins on in-dwelling devices or in wounds, to block protein-
XX CC mediated mammalian cell invasion, and to block the normal progression of
XX CC pathogenesis in infections initiated other than by the implantation of in
XX CC -dwelling devices or other surgical techniques
XX
XX Sequence 1117 BP; 312 A; 224 C; 228 G; 353 T; 0 U; 0 Other;
SQ
Query Match 76.0%; Score 19; DB 2; Length 1117;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AAAAAAAAACTATAGCTTG 20
Db 1014 AAAAAAAAACTATAGCTTG 1032

```

```

RESULT 22
AAT98574
ID AAT98574 standard; DNA; 1117 BP.
XX
XX AAT98574;
AC
XX
XX 06-NOV-1998 (first entry)
DT
XX
XX DNA encoding a S. pneumoniae ISL2 protein.
DE
XX
XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
XX KM immunological response; inoculation; antibody production; inhibitor;
XX KM T cell immune response; antimicrobial compound; bacterial adhesion;
XX KM extracellular matrix protein; protein mediated cell invasion; wound;
XX KM pathogenesis; ss.
XX
XX Streptococcus pneumoniae.
OS
XX
XX Key Location/Qualifiers
XX CDS 788..1021
XX /*tag= a
XX
XX WO9743303-A1.
XX

```

```

RESULT 23
AAK90319
ID AAK90319 standard; DNA; 460 BP.
XX
XX AAK90319;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human digestive system antigen genomic sequence SEQ ID NO: 3895.
DE
XX
XX Human, digestive system antigen; gene therapy; cancer; appendicitis;
XX KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX KM digestive system disorder; Meckel's diverticulum; ds.
XX
XX Homo sapiens.
OS
XX
XX MO200155314-A2.
XX PN
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001324.
XX PF
XX 31-JAN-2000; 2000US-0179065P.
XX

```


PS Disclosure; SEQ ID NO 3895; 986bp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 460 BP; 171 A; 71 C; 60 G; 158 T; 0 U; 0 Other;
Query Match 75.2%; Score 18.8; DB 4; Length 460;
Best Local Similarity 90.9%; Pred. No. 9.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAGCTTGAT 22
Db 210 AAAAAAAAACTAGCTTTAT 231
RESULT 24
ID AA157693 standard; DNA; 460 BP.
AC AA157693;
DT 19-OCT-2001 (first entry)
XX
XX Human colorectal cancer antigen coding sequence SEQ ID NO: 230.
XX
XX Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155350-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001350.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0198747P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 01-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.

```
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457727/49.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the colon and rectum including colorectal cancers
PT and also for testing and detection e.g. diagnosis.
XX
XX Disclosure; SEQ ID NO 230; 522pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of colorectal cancer antigens. These are shown in A157547-
CC A157619 and AAM38569-AAM38641. These can be used in the diagnosis,
CC prevention and treatment of cancer of the colon and/or rectum. The
CC present sequence is a colorectal cancer antigen genomic sequence. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 460 BP; 171 A; 71 C; 60 G; 158 T; 0 U; 0 Other;
SQ
Query Match 75.2%; Score 18.8; DB 4; Length 460;
Best Local Similarity 90.9%; Pred. No. 9.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAGCTAAGCTTAT 22
DB 210 AAAAAAAAAAAGCTAAGCTTAT 231
RESULT 25
ABS99870
ID ABS99870 standard; DNA; 460 BP.
XX
XX ABS99870;
AC
XX
XX 18-DEC-2002 (first entry)
XX
XX Genomic DNA #74 encoding human colorectal cancer related protein.
```

```
XX
XX Human, colorectal cancer related protein; colon; rectum;
KM colorectal cancer metastasis; gastrointestinal disorder; cytostatic;
KM gene; ds.
XX
XX Homo sapiens.
XX
XX US2002119919-A1.
XX
XX 29-AUG-2002.
XX
XX 17-JAN-2001; 2001US-00764855.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
PI
XX WPI; 2002-731367/79.
XX
XX New colorectal cancer polypeptide for diagnosing, prognosing, preventing,
PT and treating immune, hyperproliferative, liver, kidney, reproductive
PT disorders and for identifying modulators of therapeutic use.
XX
XX Disclosure; SEQ ID NO 230; 183pp; English.
XX
XX The present invention relates to the isolation of novel human colorectal
CC cancer related proteins, and polynucleotide sequences encoding them. The
CC sequences of the invention are useful in the diagnosis, treatment,
CC prevention and/or prognosis of the colon and/or rectum, including
CC colorectal cancer, colorectal cancer metastases, and gastrointestinal
CC disorders such as dysphagia, peptic oesophagitis, gastric reflux,
CC irritable bowel syndrome, and peritoneal diseases. The invention also
CC describes antibodies that bind colorectal cancer related proteins,
CC vectors, host cells, and recombinant and synthetic methods for producing
CC human colorectal cancer related polynucleotides, polypeptides, and/or
CC antibodies. ABS99797-ABS99974 represent genomic sequences encoding human
CC colorectal cancer related proteins. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site at
CC seqdata.uspto.gov/psipdidentry.html
XX
XX Sequence 460 BP; 171 A; 71 C; 60 G; 158 T; 0 U; 0 Other;
SQ
Query Match 75.2%; Score 18.8; DB 6; Length 460;
Best Local Similarity 90.9%; Pred. No. 9.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAGCTAAGCTTAT 22
DB 210 AAAAAAAAAAAGCTAAGCTTAT 231
RESULT 26
ADB93023
ID ADB93023 standard; DNA; 460 BP.
XX
XX ADB93023;
AC
XX
XX 04-DEC-2003 (first entry)
XX
XX Human colorectal cancer related polypeptide DNA #74.
XX
XX ds; gene; human; colorectal cancer; antigen; gene therapy;
KM gastrointestinal disorder; inflammatory disease; infection; cancer;
KM intestinal neoplasm; small intestine carcinoma tumour;
KM small intestine non-Hodgkin's lymphoma; small bowel lymphoma; ulcer;
KM peptic ulcer; Bruton's disease; X linked infantile agammaglobulinemia;
KM severe combined immunodeficiency; DisGeorge anomaly;
KM hyperproliferative disorder; acute lymphoblastic leukaemia;
```

KM acute lymphocytic leukaemia; urinary system disorder; cortical necrosis;
KM kidney infection; cardiovascular disorder; carcinoid heart disease;
KM arrhythmia; respiratory disorder; non-allergic rhinitis; sinusitis;
KM musculoskeletal system disorder; Albers-Schönberg disease;
KM Marfan's syndrome; neurological disease; phenylketonuria;
KM Merrick's encephalopathy; Alzheimer's disease; endocrine disorder;
KM Grave's disease; Cushing's syndrome; reproductive system disorder;
KM prostaticositis; benign prostatic hypertrophy; benign prostatic hyperplasia;
KM thrombosis; atherosclerosis; myocardial infarction; ischaemic attack.
OS Homo sapiens.
XX
XX
PN US2003054420-A1.
PD 20-MAR-2003.
XX
PF 11-FEB-2002; 2002US-00072349.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231422P.
PR 08-SEP-2000; 2000US-0231433P.
PR 08-SEP-2000; 2000US-0231444P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 14-SEP-2000; 2000US-0233066P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0234999P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246539P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.

PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241212P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246509P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251088P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-502630/55.
XX
XX P-PSDB; AAM82609.
XX
XX Polynucleotides encoding digestive system antigens, useful for
XX PT diagnosing, treating, preventing and/or prognosing disorders of the
XX PT digestive system, particularly cancer and cancer metastases.
XX

PS Claim 1; SEQ ID NO 698; 986bp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC system antigen of the invention
XX
SQ Sequence 469 BP; 173 A; 71 C; 60 G; 155 T; 0 U; 10 Other;
XX
Query Match 75.2%; Score 18.8; DB 4; Length 469;
Best Local Similarity 90.9%; Pred. No. 9.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTTGAT 22
DB 211 AAAAAAAAAAGCTTAGCTTAT 232
RESULT 28
AA157594
ID AA157594 strand: cDNA; 469 BP.
XX
AC AA157594;
XX
DT 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen cDNA SEQ ID NO: 58.
XX
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN MO200155350-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001350.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.

DE cDNA encoding human colorectal cancer related protein #48.
XX
XX Human; colorectal cancer related protein; colon; rectum;
KW colorectal cancer metastasis; gastrointestinal disorder; cytostatic;
KW gene; ss.
XX
XX Homo sapiens.
XX
XX US2002119919-A1.
XX
XX 29-AUG-2002.
XX
XX
XX 17-JAN-2001; 2001US-00764855.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
PI
XX MPI; 2002-731367/79.
XX
XX P-PSDB; ABG97668.
XX
XX New colorectal cancer polypeptide for diagnosing, prognosing, preventing,
PT and treating immune, hyperproliferative, liver, kidney, reproductive
PT disorders and for identifying modulators of therapeutic use.
XX
XX Claim 4; SEQ ID NO 58; 183bp; English.
XX
XX The present invention relates to the isolation of novel human colorectal
XX cancer related proteins, and polynucleotide sequences encoding them. The
XX sequences of the invention are useful in the diagnosis, treatment,
XX prevention and/or prognosis of the colon and/or rectum, including
XX colorectal cancer. colorectal cancer metastases, and gastrointestinal
XX disorders such as dysphagia, peptic esophagitis, gastric reflux,
XX irritable bowel syndrome, and peritoneal diseases. The invention also
XX describes antibodies that bind colorectal cancer related proteins,
XX vectors, host cells, and recombinant and synthetic methods for producing
XX human colorectal cancer related polynucleotides, polypeptides, and/or
XX antibodies. AB899724-AB899796 represent cDNA sequences encoding human
XX colorectal cancer related proteins. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from the USPTO web site at
XX seqdata.uspto.gov/psipdsidentry.html
XX
XX Sequence 469 BP; 173 A; 71 C; 60 G; 155 T; 0 U; 10 Other;
SQ
Query Match 75.2%; Score 18.8; DB 6; Length 469;
Best Local Similarity 90.3%; Pred. No. 9.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAGCTTAT 22
Db 211 AAAAAAAAAAGCTATAGCTTAT 232
RESULT 30
ADB92851
ID ADB92851 standard; cDNA; 469 BP.
XX
XX ADB92851;
XX
XX 04-DEC-2003 (first entry)
XX
XX Human colorectal cancer related polypeptide cDNA #48.
XX
XX ss; gene; human; colorectal cancer; antigen; gene therapy;
KW gastrointestinal disorder; inflammatory disease; infection; cancer;
KW intestinal neoplasm; small intestine carcinoma tumor;
KW small intestine non-Hodgkin's lymphoma; small bowel lymphoma; ulcer;
KW peptic ulcer; Bruton's disease; X linked infantile agammaglobulinemia;

KW severe combined immunodeficiency; DiGeorge anomaly;
KW hyperproliferative disorder; acute lymphoblastic leukemia;
KW acute lymphocytic leukemia; urinary system disorder; cortical necrosis;
KW kidney infection; cardiovascular disorder; carcinoma heart disease;
KW arrhythmia; respiratory disorder; non-allergic rhinitis; sinusitis;
KW musculoskeletal system disorder; Albers-schönberg disease;
KW Marfan's syndrome; neurological disease; phenylketonuria;
KW Wernicke's encephalopathy; Alzheimer's disease; endocrine disorder;
KW Grave's disease; Cushing's syndrome; reproductive system disorder;
KW prostatitis; benign prostatic hyper trophy; benign prostatic hyperplasia;
KW thrombosis; atherosclerosis; myocardial infarction; ischemic attack.
XX
XX Homo sapiens.
XX
XX US2003054420-A1.
XX
XX 20-MAR-2003.
XX
XX 11-FEB-2002; 2002US-00072349.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205115P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 07-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 14-JUL-2000; 2000US-0218290P.
XX
XX 26-JUL-2000; 2000US-0220963P.
XX
XX 26-JUL-2000; 2000US-0220964P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
XX
XX 14-AUG-2000; 2000US-0225270P.
XX
XX 14-AUG-2000; 2000US-0225447P.
XX
XX 14-AUG-2000; 2000US-0225757P.
XX
XX 14-AUG-2000; 2000US-0225758P.
XX
XX 14-AUG-2000; 2000US-0225759P.
XX
XX 18-AUG-2000; 2000US-0226279P.
XX
XX 22-AUG-2000; 2000US-0226681P.
XX
XX 22-AUG-2000; 2000US-0226682P.
XX
XX 22-AUG-2000; 2000US-0227182P.
XX
XX 23-AUG-2000; 2000US-0227009P.
XX
XX 30-AUG-2000; 2000US-0228924P.
XX
XX 01-SEP-2000; 2000US-0229287P.
XX
XX 01-SEP-2000; 2000US-0229343P.
XX
XX 01-SEP-2000; 2000US-0229344P.
XX
XX 01-SEP-2000; 2000US-0229345P.
XX
XX 05-SEP-2000; 2000US-0229509P.
XX
XX 05-SEP-2000; 2000US-0229513P.
XX
XX 06-SEP-2000; 2000US-0230437P.
XX
XX 06-SEP-2000; 2000US-0230438P.
XX
XX 08-SEP-2000; 2000US-0231242P.
XX
XX 08-SEP-2000; 2000US-0231243P.
XX
XX 08-SEP-2000; 2000US-0231244P.
XX
XX 08-SEP-2000; 2000US-0231413P.
XX
XX 08-SEP-2000; 2000US-0231414P.
XX
XX 08-SEP-2000; 2000US-0232080P.
XX
XX 08-SEP-2000; 2000US-0232081P.
XX
XX 12-SEP-2000; 2000US-0231968P.
XX
XX 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232389P.
 PR 14-SEP-2000; 2000US-0232389P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234224P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236377P.
 PR 29-SEP-2000; 2000US-0236379P.
 PR 29-SEP-2000; 2000US-0236386P.
 PR 29-SEP-2000; 2000US-0236389P.
 PR 29-SEP-2000; 2000US-0236392P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241865P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246539P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249219P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249247P.
 PR 17-NOV-2000; 2000US-0249248P.
 PR 17-NOV-2000; 2000US-0249249P.
 PR 17-NOV-2000; 2000US-0249250P.
 PR 17-NOV-2000; 2000US-0249251P.
 PR 17-NOV-2000; 2000US-0249252P.
 PR 17-NOV-2000; 2000US-0249253P.
 PR 17-NOV-2000; 2000US-0249254P.
 PR 17-NOV-2000; 2000US-0249255P.
 PR 17-NOV-2000; 2000US-0249256P.
 PR 17-NOV-2000; 2000US-0249257P.
 PR 17-NOV-2000; 2000US-0249258P.
 PR 17-NOV-2000; 2000US-0249259P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR 17-JAN-2001; 2001US-00764855.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2003-708345/67.
 DR P-PSDB; ADB92924.
 XX
 PT Novel colorectal cancer antigen useful for treating, preventing,
 PT diagnosing and/or prognosing gastrointestinal disorders, infections,
 PT cancers such as intestinal neoplasms, ulcers.
 XX
 PS Claim 3; SEQ ID NO 58; 179pp; English.
 XX
 CC The invention relates to a colorectal cancer antigen. The antigen is
 CC useful for chromosome identification, chromosome mapping, radiation
 CC hybrid mapping or gene therapy, or as hybridisation probes for
 CC differential identification of the tissues or cell types present in a
 CC biological sample. The antigen is useful for treating, preventing,
 CC diagnosing and/or prognosing gastrointestinal disorders, including
 CC inflammatory diseases and/or conditions, infections, cancers (e.g.
 CC intestinal neoplasms (carcinoid tumour of the small intestine, non-
 CC Hodgkin's lymphoma of the small intestine, small bowel lymphoma)) and
 CC ulcers (e.g. peptic ulcers). The antigen and its nucleic acid are useful
 CC to provide immunological probes for differential identification of the
 CC tissue. The antigen and its nucleic acid are useful for treating,
 CC preventing, diagnosing and/or prognosing diseases, disorders and/or
 CC conditions of the immune system e.g. Bruton's disease, X linked infantile
 CC agammaglobulinaemia, severe combined immunodeficiencies, DiGeorge
 CC anomaly, etc. The antigen and its nucleic acid is useful for treating,
 CC preventing and/or diagnosing hyperproliferative disorders (e.g. acute
 CC
 Query Match 75.2%; Score 18.8; DB 10; Length 469;
 Best Local Similarity 90.9%; Pred. No. 9.9e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTATGCTTAT 22
 Db 211 AAAAAAAAACTATGCTTAT 232
 ACN61173/c
 ID ACN61173 standard; cDNA, 573 BP.
 AC ACN61173;
 XX
 AC ACN61173;
 AC
 DT 02-DEC-2004 (first entry)
 XX
 DE Cotton gynoecium tissue EST Clone ID: LIB3829-032-06-NE-G12, SEQ.15954.
 XX
 KW Cotton; plant; EST; expressed sequence tag; transgenic plant; gynoecium;
 KW variety Nucleon3B; library LIB3829; molecular tag; molecular marker;
 KW genetic mapping; molecular mapping; seed germination; plant growth;
 KW plant quality; plant yield; plant breeding; tissue printing; ss.
 XX
 OS Gossypium hirsutum.
 OS
 XX
 PN US2004123340-A1.
 XX
 PD 24-JUN-2004.
 XX
 PF 12-DEC-2001; 2001US-00021323.

```

XX 14-DEC-2000; 2000US-025619P.
PR
XX
XX (DEIK/) DEIRMAN J.
PA (FENG/) FENG P C C.
PA (FINC/) FINCHER K L.
PA (ZIEG/) ZIEGLER T E.
PI Delkman J, Feng PCC, Fincher KL, Ziegler TE;
XX WPI; 2004-479808/45.
XX
XX New isolated nucleic acid molecule that encodes a plant protein or its
XX fragment, useful for isolating a variety of agronomically significant
XX genes associated with plant growth, quality or yield, and as molecular
XX tags to map genes.
XX
XX Claim 1: SEQ ID NO 15954; 34pp; English.
XX
XX The invention relates to 17880 cotton expressed sequence tags (ESTs;
XX ACN45220-ACN63093). The ESTs were isolated from cDNA libraries generated
XX from primed or non-primed seeds from variety DP50B, mature seeds from
XX variety Coker 312, Boswell 96 Field, and androecium tissue, gynoecium
XX tissue, developing fibres, carpel walls and septa from variety
XX Nucleoton33B. The invention also relates to substantially purified
XX proteins or their fragments encoded by nucleic acid molecules of the
XX invention, and to transformed plants having a nucleic acid construct
XX comprising a nucleic acid of the invention. The cotton ESTs are useful as
XX molecular tags to isolate genetic regions, to isolate genes, to map
XX genes, to determine gene function and to determining whether genes are
XX members of a particular gene family. The nucleic acid molecules may be
XX used for isolating a variety of agronomically significant genes
XX associated with plant growth, quality, yield, and could also serve as
XX links in metabolic and catabolic pathways. The nucleic acid molecules are
XX also useful for identifying genes important in initiating and maintaining
XX seed germination or that may be used to mitigate stresses encountered
XX during seed germination. The ESTs additionally enable the acquisition of
XX promoters and cis-regulatory elements which will be useful to express
XX agronomically significant genes in these tissues and/or other tissues,
XX and also permits the acquisition of molecular markers useful in breeding
XX schemes, genetic and molecular mapping, and in cloning of agronomically
XX significant genes. The nucleic acid molecules are further useful for
XX detecting the expression level or pattern of a protein or mRNA and for
XX detecting the presence or quantity of a protein by tissue printing. The
XX present sequence represents a specifically claimed EST isolated from a
XX cotton variety Nucleoton33B gynoecium tissue cDNA library (L1B3829). The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the US
XX patent office at seqdata.uspto.gov/sequence.html?docID=US20040123340
XX
XX Sequence 573 BP; 208 A; 112 C; 90 G; 163 T; 0 U; 0 Other;
SQ
XX
XX Query Match 75.2%; Score 18.8; DB 13; Length 573;
XX Best Local Similarity 90.9%; Pred. No. 9.9e+02;
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAAGTATGCTTGAT 22
XX |||||
XX DB 180 AAAAAAAAAAAGTATGCTTGAT 159
XX
XX RESULT 32
XX AAS46525/C
XX ID AAS46525 standard; DNA; 6749 BP.
XX
XX AAS46525;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #247.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX

```

```

XX cytosine methylaction; ds.
XX
XX Homo sapiens.
XX
XX MO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX
XX 15-MAR-2001; 2001WO-EP002955.
XX
XX
XX 15-MAR-2000; 2000DE-01013847.
XX PR 06-APR-2000; 2000DE-01019058.
XX PR 07-APR-2000; 2000DE-01019173.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumor suppressor
XX genes and oncogenes, useful in designing primers and probes for analyzing
XX diseases associated with cytosine methylation state e.g. cancer.
XX
XX Claim 1: SEQ ID NO 247; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (cp DNA) e.g. with
XX bisulphite, of genes associated with tumor suppression and oncogenes
XX having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX 500 are missing from the sequence listing) sequences (Ss) and sequences
XX complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
XX oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX probes for detecting the cytosine methylation state and/or single
XX nucleotide polymorphisms and also to be used in an array for analysing
XX diseases associated with CpG dinucleotides e.g. cancers and tumours. The
XX probes can also be used in a method for ascertaining genetic and/or
XX epigenetic parameters for the diagnosis and/or therapy of existing
XX diseases or the predisposition to specific diseases, by analysing
XX cytosine methylations. The parameters may be compared to another set of
XX genetic and/or epigenetic parameters, the differences serving as basis
XX for diagnosis and/or prognosis events which are disadvantageous to
XX patients. The present sequence is one of the 533 genomic sequences
XX derived from tumour suppressor genes and oncogenes. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published\_pct\_sequences
XX
XX Sequence 6749 BP; 1953 A; 172 C; 1640 G; 2984 T; 0 U; 0 Other;
SQ
XX
XX Query Match 75.2%; Score 18.8; DB 4; Length 6749;
XX Best Local Similarity 90.9%; Pred. No. 1.1e+03;
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAAGTATGCTTGAT 22
XX |||||
XX DB 3176 AAAAAAAAAAAGTATGCTTGAT 3155
XX
XX RESULT 33
XX AAS46336/C
XX ID AAS46336 standard; DNA; 7348 BP.
XX
XX AAS46336;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #58.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX

```

KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN MO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 58; 27pp; English.
XX
SQ The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (58) and sequences
CC complementary to (58). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seg ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7348 BP; 2162 A; 37 C; 1405 G; 3744 T; 0 U; 0 Other;
Query Match 75.2%; Score 18.8; DB 4; Length 7348;
Best Local Similarity 90.9%; Pred. No. 1.le+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTATAGCTTAT 22
DB 280 AAAAAAAAACTATAGCTTAT 259
RESULT 34
ABL33435/c
ID ABL33435 standard; DNA; 7703 BP.
XX
XX ABL33435;
AC
XX
DT 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1408.

XX
KW Human; immune system disease; cytosine methylation; antiaethmatic;
KW antiarteriosclerotic; antihaemic; cyostatic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.
XX
OS Homo sapiens.
XX
PN MO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1408; 32pp + Sequence Listing; German.
XX
SQ The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 7703 BP; 1808 A; 87 C; 1841 G; 3967 T; 0 U; 0 Other;
Query Match 75.2%; Score 18.8; DB 6; Length 7703;
Best Local Similarity 90.9%; Pred. No. 1.le+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 AAAAAAAAACTATAGCTTATCT 24
DB 6575 AAAAAAAAACTATAGCTTATCT 6554
RESULT 35
ABL33323/c
ID ABL33323 standard; DNA; 9741 BP.
XX
XX ABL33323;
AC
XX
DT 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1296.
DE
XX
KW Human; immune system disease; cytosine methylation; antiaethmatic;
KW antiarteriosclerotic; antihaemic; cyostatic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.
XX
OS Homo sapiens.

```
XX PN WO200200928-A2.
XX PD
XX PF 03-JAN-2002.
XX PR
XX PS 02-JUL-2001; 2001WO-BP007537.
XX PT
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PA (EPiG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
XX PT diagnosis and treatment of diseases associated with abnormal cytosine
XX PT methylation.
XX PS
XX PS Claim 1, SEQ ID NO 1296; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention
XX SQ
XX SQ Sequence 9741 BP; 2623 A; 195 C; 1863 G; 5060 T; 0 U; 0 Other;
XX
XX Query Match 75.2%; Score 18.8; DB 6; Length 9741;
XX Best Local Similarity 90.9%; Pred. No. 1.1e+03;
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAACTATAGCTTGAT 22
XX DB 1010 AAAAAAAAACTATAGCTTTAT 989
XX
XX RESULT 36
XX ID AAT91324 standard; DNA; 12987 BP.
XX XX
XX AC AAT91324;
XX XX
XX DT 27-APR-1998 (first entry)
XX XX
XX DE Arabidopsis thaliana inorganic phosphate transporter 3 and 2 DNA.
XX XX
XX KM Columbia strain; inorganic phosphate transporter 3; IP13;
XX KM inorganic phosphate transporter 2; IP12; accelerated phosphate uptake;
XX KM tobacco plant; ds.
XX OS Arabidopsis thaliana.
XX OS
XX FH Key Location/Qualifiers
XX FT CDS 2561..4251
XX FT /tag= a
XX FT /product= "inorganic_phosphate_transporter_3"
XX FT /tag= c
XX FT 2962..3086
XX FT /tag= c
XX FT /number= 1
XX FT intron 4249..9110
XX FT /tag= d
XX FT /number= 2
XX FT CDS 9111..10835
XX FT /tag= b
XX FT /product= "inorganic_phosphate_transporter_2"
XX PN WO9735984-A1.
```

```
XX PD 02-OCT-1997.
XX XX
XX PF 24-MAR-1997; 97WO-JP000975.
XX PR
XX PR 25-MAR-1996; 96JP-00094790.
XX XX
XX PA (MITS-) MITSUI PLANT BIOTECHNOLOGY RES INST.
XX PI Mitukawa N, Okumura S, Shirano Y, Shibata D;
XX DR WPI; 1997-489647/45.
XX DR P-PSDB; AAM32303, AAM32304.
XX XX
XX PT DNA encoding a plant phosphate transporter protein - useful for producing
XX PT e.g. tobacco plants with increased phosphate uptake and accelerated
XX PT growth.
XX PS
XX PS Example 3, Page 40-54; 96pp; Japanese.
XX CC The present sequence encodes Arabidopsis thaliana (columbia strain)
XX CC inorganic phosphate transporters 3 and 2 (IP13 and 2), useful for
XX CC accelerating the uptake of phosphate by plants, e.g. tobacco plants
XX SQ
XX SQ Sequence 12987 BP; 4319 A; 2171 C; 2173 G; 4324 T; 0 U; 0 Other;
XX
XX Query Match 75.2%; Score 18.8; DB 2; Length 12987;
XX Best Local Similarity 90.9%; Pred. No. 1.1e+03;
XX Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 3 AAAAAAAAACTATAGCTTGATCT 24
XX DB 3520 AAAAAAAAACTATAGCTTGTTCT 3541
XX
XX RESULT 37
XX ID ACN37235 standard; DNA; 47038 BP.
XX AC ACN37235;
XX XX
XX DT 18-NOV-2004 (first entry)
XX XX
XX DE Human periodontal disease related gene BQFR SEQ ID NO:145.
XX XX
XX KM periodontal disease; polymorphism; ds; human; gene; SNP;
XX KM single nucleotide polymorphism.
XX OS Homo sapiens.
XX OS
XX FH Key Location/Qualifiers
XX FT misc_feature 75
XX FT /tag= a
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT /note= "Variable nucleotide C,T"
XX FT 2800
XX FT /tag= b
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT /note= "Variable nucleotide A,G"
XX FT 3010
XX FT /tag= c
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT /note= "Variable nucleotide A,G"
XX FT misc_feature 3957
XX FT /tag= d
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT /note= "Variable nucleotide C,G"
XX FT 4627
XX FT /tag= e
XX FT /standard_name= "Single nucleotide polymorphism"
XX FT /note= "Variable nucleotide C,T"
XX FT 4722
XX FT /tag= f
XX FT misc_feature
```

FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide C,T"
FT	/note= "Variable nucleotide G,T"	FT	11329
FT	misc_feature	FT	/tag= y
FT	4825	FT	/standard_name= "Single nucleotide polymorphism"
FT	/tag= g	FT	/note= "Variable nucleotide C,T"
FT	/standard_name= "Single nucleotide polymorphism"	FT	11615
FT	/note= "Variable nucleotide C,T"	FT	/tag= z
FT	4881	FT	/standard_name= "Single nucleotide polymorphism"
FT	/tag= h	FT	/note= "Variable nucleotide A,G"
FT	/standard_name= "Single nucleotide polymorphism"	FT	12098
FT	/note= "Variable nucleotide A,G"	FT	/tag= aa
FT	5253	FT	/standard_name= "Single nucleotide polymorphism"
FT	misc_feature	FT	/note= "Variable nucleotide A,G"
FT	7319	FT	13256
FT	/tag= i	FT	/tag= ab
FT	/standard_name= "Single nucleotide polymorphism"	FT	/standard_name= "Single nucleotide polymorphism"
FT	/note= "Variable nucleotide C,T"	FT	/note= "Variable nucleotide C,T"
FT	7319	FT	13279
FT	/tag= j	FT	/tag= ac
FT	/standard_name= "Single nucleotide polymorphism"	FT	/standard_name= "Single nucleotide polymorphism"
FT	/note= "Variable nucleotide A,G"	FT	/note= "Variable nucleotide C,T"
FT	8087	FT	15458
FT	/tag= k	FT	/tag= ad
FT	/standard_name= "Single nucleotide polymorphism"	FT	/standard_name= "Single nucleotide polymorphism"
FT	/note= "Variable nucleotide C,T"	FT	/note= "Variable nucleotide C,T"
FT	8396	FT	17422
FT	misc_feature	FT	/tag= ae
FT	/tag= l	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide G,T"
FT	/note= "Variable nucleotide T,C"	FT	18635
FT	8834	FT	/tag= af
FT	/tag= m	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide C,A"
FT	/note= "Variable nucleotide A,G"	FT	19920
FT	9100	FT	/tag= ag
FT	/tag= n	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide T,C"
FT	/note= "Variable nucleotide C,T"	FT	23233
FT	9105	FT	/tag= ah
FT	/tag= o	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide C,G"
FT	/note= "Variable nucleotide C,G"	FT	23710
FT	9826	FT	/tag= ai
FT	/tag= p	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide C,T"
FT	/note= "Variable nucleotide A,T"	FT	24050
FT	9835	FT	/tag= aj
FT	/tag= q	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide T,C"
FT	/note= "Variable nucleotide A,G"	FT	24751
FT	10001	FT	/tag= ak
FT	/tag= r	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide G,T"
FT	/note= "Variable nucleotide C,T"	FT	25417
FT	10070	FT	/tag= al
FT	/tag= s	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide C,T"
FT	/note= "Variable nucleotide A,G"	FT	26209
FT	10598	FT	/tag= am
FT	/tag= t	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide C,T"
FT	/note= "Variable nucleotide A,C"	FT	26441
FT	10618	FT	/tag= an
FT	/tag= u	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide A,T"
FT	/note= "Variable nucleotide A,G"	FT	28826
FT	10631	FT	/tag= ao
FT	/tag= v	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide C,T"
FT	/note= "Variable nucleotide A,T"	FT	29255
FT	10642	FT	/tag= ap
FT	/tag= w	FT	/standard_name= "Single nucleotide polymorphism"
FT	/standard_name= "Single nucleotide polymorphism"	FT	/note= "Variable nucleotide A,C"
FT	/note= "Variable nucleotide A,T"	FT	
FT	10718	FT	
FT	misc_feature	FT	
FT	/tag= x	FT	
FT	/standard_name= "Single nucleotide polymorphism"	FT	

```
FT misc_feature 29296
FT /tag= aq
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide A,G"
FT misc_feature 30119
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide A,G"
FT misc_feature 31377
FT /tag= as
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide T,C"
FT misc_feature 31512
FT /tag= at
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Variable nucleotide A,G"
FT misc_feature 31699
FT /tag= au

Query Match 75.2%; Score 18.8; DB 13; Length 47038;
Best Local Similarity 90.9%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAGCTTGAT 22
Db 40236 AAAAAAACTTAGCTTTAT 40257

RESULT 38
ABD32854
ID ABD32854 standard; DNA; 209083 BP.
XX
AC ABD32854;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD17-032.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
XX leukemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
XX 14-MAR-2003; 2003US-00388838.
XX 15-APR-2003; 2003US-00417375.
XX 13-JUN-2003; 2003US-00461862.
XX 15-SEP-2003; 2003US-00663431.
XX 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
XX for diagnosing, preventing or treating cancers, especially lymphoma and
XX leukemia, or in screening for agents that modulate cancer.
XX
PS claim 16; seqid 524; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
```

```
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells (comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 209083 BP; 57124 A; 47057 C; 45679 G; 59223 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 13; Length 209083;
Best Local Similarity 90.9%; Pred. No. 1.1e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAGCTTGAT 22
Db 206237 AAAAAAACTTAGCTTTAT 206258

RESULT 39
ABV06607/C
ID ABV06607 standard; cDNA; 312 BP.
XX
AC ABV06607;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 6598.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189862P.
XX 25-MAY-2000; 2000US-0207454P.
XX 09-JUN-2000; 2000US-0211314P.
XX 18-JUL-2000; 2000US-0219007P.
XX 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
```

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

XX

PS Claim 1; Page 1079-1080; 11750pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the CC specification or its complement. (I) is useful for: (a) assessing whether CC a patient is afflicted with prostate cancer; (b) monitoring the CC progression of prostate cancer in a patient; (c) assessing the efficacy CC of a test compound to inhibit prostate cancer in a patient; (d) assessing CC the efficacy of a therapy for inhibiting prostate cancer in a patient; CC (e) selecting a composition for inhibiting prostate cancer in a patient; CC (f) assessing the prostate cell carcinogenic potential of a compound; (g) CC determining whether prostate cancer has metastasized in a patient; (h) CC assessing the aggressiveness or indolence of prostate cancer in a patient CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker XX

SQ Sequence 312 BP; 67 A; 66 C; 63 G; 115 T; 0 U; 1 Other;

Query Match 74.4%; Score 18.6; DB 5; Length 312;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTATCTT 25
DB 140 AAAAAAAAAAGTACGCTTCTT 116

RESULT 40
ABV36556/c
ID ABV36556 standard; cDNA; 391 BP.

XX

AC ABV36556;

XX

DT 16-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 36547.

XX

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US005171.

XX

PR 17-FEB-2000; 2000US-0183319P.

XX

PR 16-MAR-2000; 2000US-0189862P.

XX

PR 25-MAY-2000; 2000US-0207454P.

XX

PR 09-JUN-2000; 2000US-0211314P.

XX

PR 18-JUL-2000; 2000US-0219007P.

XX

PR 13-DEC-2000; 2000US-0255281P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Schlegel R, Endege WO, Monahan JE;

XX

DR WPI; 2001-662795/76.

XX

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

XX

PS Claim 1; Page 7549; 11750pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the CC progression of prostate cancer in a patient; (c) assessing the efficacy CC of a test compound to inhibit prostate cancer in a patient; (d) assessing CC the efficacy of a therapy for inhibiting prostate cancer in a patient; CC (e) selecting a composition for inhibiting prostate cancer in a patient; CC (f) assessing the prostate cell carcinogenic potential of a compound; (g) CC determining whether prostate cancer has metastasized in a patient; (h) CC assessing the aggressiveness or indolence of prostate cancer in a patient CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker XX

SQ Sequence 391 BP; 85 A; 86 C; 88 G; 132 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 5; Length 391;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTATCTT 25
DB 220 AAAAAAAAAAGTACGCTTCTT 196

RESULT 41
ABV04081/c
ID ABV04081 standard; cDNA; 424 BP.

XX

AC ABV04081;

XX

DT 13-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 4072.

XX

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US005171.

XX

PR 17-FEB-2000; 2000US-0183319P.

XX

PR 16-MAR-2000; 2000US-0189862P.

XX

PR 25-MAY-2000; 2000US-0207454P.

XX

PR 09-JUN-2000; 2000US-0211314P.

XX

PR 18-JUL-2000; 2000US-0219007P.

XX

PR 13-DEC-2000; 2000US-0255281P.

XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI Schlegel R, Endege WO, Monahan JE;

XX

DR WPI; 2001-662795/76.

XX

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

XX

PS Claim 1; Page 716; 11750pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the CC specification or its complement. (I) is useful for: (a) assessing whether CC a patient is afflicted with prostate cancer; (b) monitoring the CC progression of prostate cancer in a patient; (c) assessing the efficacy CC of a test compound to inhibit prostate cancer in a patient; (d) assessing CC the efficacy of a therapy for inhibiting prostate cancer in a patient; CC (e) selecting a composition for inhibiting prostate cancer in a patient; CC (f) assessing the prostate cell carcinogenic potential of a compound; (g) CC determining whether prostate cancer has metastasized in a patient; (h) CC assessing the aggressiveness or indolence of prostate cancer in a patient CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX SQ Sequence 424 BP; 101 A; 99 C; 88 G; 133 T; 0 U; 3 Other;
 Query Match 74.4%; Score 18.6; DB 5; Length 424;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAAGTTAGCTTGATCTT 25
 DB 374 AAAAAAAAAAAGTTAGCTTGATCTT 350

RESULT 42
 ABRN21922/c
 ID ABRN21922 standard; cDNA; 433 BP.
 XX AC ABRN21922;
 XX DT 24-JUN-2002 (first entry)
 XX DE Human ORFX polynucleotide sequence SEQ ID NO:12321.
 XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 XX KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 XX KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 XX KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 XX KW hypertension; hypothyroidism; cholesterol ester storage disease;
 XX KW immune deficiency; immune disorder; infectious disease;
 XX KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 XX KW myasthenia gravis; gene; ss.
 XX OS Homo sapiens.
 XX PN WO200192523-A2.
 XX PD 06-DEC-2001.
 XX PP 29-MAY-2001; 2001WO-US010836.
 XX PR 30-MAY-2000; 2000US-0206132P.
 XX PR 29-AUG-2000; 2000US-0228716P.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Shimkete RA, Leach MD;
 XX PI P-PSDB; ABR06170.
 XX DR WPI; 2002-106308/14.
 XX DR P-PSDB; ABR06170.
 XX PT Novel human polypeptides and polynucleotides useful for diagnosing,
 XX PT preventing and treating cardiovascular disease, neurodegenerative,
 XX PT hyperproliferative disorders and autoimmune disorders.
 XX PS Disclosure; SEQ ID NO 12321; 1037pp; English.

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see table 1 in the specification). ABRN15762 to ABRN2752 encode the human ORFX proteins given in ABR00010 to ABR1500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis.

CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WFO at ftp.wfo.int/pub/published_pct_sequences
 XX SQ Sequence 433 BP; 97 A; 104 C; 85 G; 146 T; 0 U; 1 Other;
 Query Match 74.4%; Score 18.6; DB 6; Length 433;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAAGTTAGCTTGATCTT 25
 DB 194 AAAAAAAAAAAGTTAGCTTGATCTT 170

RESULT 43
 AAV90433
 ID AAV90433 standard; cDNA; 438 BP.
 XX AC AAV90433;
 XX DT 15-FEB-1999 (first entry)
 XX DE EST clone DN144.
 XX KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
 XX KW tissue growth; activin; inhibin; chemokines; chemokines; haemostatic;
 XX KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 XX KW gene therapy; ss.
 XX OS Homo sapiens.
 XX PN WO9845436-A2.
 XX PD 15-OCT-1998.
 XX PP 10-APR-1998; 98WO-US006955.
 XX PR 10-APR-1997; 97US-00838821.
 XX PA (GENETICS INST INC.
 XX PI Jacobs K, McCoy JM, Lavallee ER, Racie LA, Werberg D, Treacy M;
 XX PI Spaulding V, Agostini MJ;
 XX DR WPI; 1999-070077/06.
 XX PT New polynucleotides encoding human secreted proteins - derived from e.g.
 XX PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,
 XX PT pituitary, retina and colon cDNA libraries.
 XX PS Claim 1; Page 539; 618pp; English.

The present sequence represents a human expressed sequence tag (EST). The CC polynucleotide, which is a secreted EST, and the encoded protein are CC predicted to have useful biological activities which would make them CC suitable for treating, preventing or ameliorating medical conditions in CC humans and animals, although no supporting data is given. Suggested CC activities include nutritional activity, immune stimulating or CC suppressing activity, haematopoiesis regulating activity, tissue growth CC activity, activin/inhibin activity, chemotactic/chemokinetic activity, CC haemostatic and thrombolytic activity, receptor/ligand activity, anti- CC inflammatory activity, cadherin/tumour invasion suppressor activity.
 CC tumour inhibition activity. The polynucleotide may also be useful for CC gene therapy

XX SQ Sequence 438 BP; 135 A; 74 C; 136 G; 93 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 2; Length 438;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTTGATCTT 25
 Db 125 AAAAAAAAAAGCTACCTTGATCTT 149

RESULT 44
 ADX53399/C
 ID ADX53399 standard; cDNA; 664 BP.
 XX
 AC ADX53399;
 XX
 DT 21-APR-2005 (first entry)
 XX

DE Plant full length insert polynucleotide seqid 28139.
 XX
 KM plant protectant; plant growth regulant; gene therapy; plant;
 KM recombinant DNA construct; physical array; plant breeding marker;
 KM cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KM extreme osmotic condition; pathogen tolerance; pest tolerance;
 KM growth rate; cell cycle pathway; disease resistance;
 KM galactomanan production; lignin production; plant growth regulator;
 KM yield; plant growth; plant development; seed oil; protein yield;
 KM protein content; gene; ss.
 XX

OS Unidentified.
 XX

PN US2004034888-A1.
 XX

PD 19-FEB-2004.
 XX

PF 28-APR-2003; 2003US-00425114.
 XX

PR 06-MAY-1999; 99US-00304517.
 XX

PR 05-NOV-2001; 2001US-00985678.
 XX

PA (LIU/) LIU J.
 XX

PA (KOV/) KOVALIC D K.
 XX

PA (SCRE/) SCREEN S E.
 XX

PA (TAB/) TABASKA J E.
 XX

PA (CAO/) CAO Y.
 XX

PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
 XX

DR MPI; 2004-180133/17.
 XX

PT New recombinant DNA construct, useful for improving plant tolerance to
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
 PT pests, for conferring increased resistance to plant disease, or for
 PT improving yield.
 XX

Claim 1; SEQ ID NO 28139; 15pp; English.
 XX

CC The invention describes a recombinant DNA construct comprising a
 CC polynucleotide consisting of a sequence encoding an amino acid sequence
 CC available in electronic form from the US patent office at
 CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
 CC of the invention are also useful in physical arrays of molecules and as
 CC plant breeding markers. The recombinant DNA construct is useful for
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in
 CC plant cells by modification of the cell cycle pathway, for conferring
 CC increased resistance to plant disease, for producing galactomanan,
 CC lignin or plant growth regulators, for increasing the rate of homologous
 CC recombination in plants, for improving yield by modification of
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This sequence represents a plant full length insert
 CC polynucleotide that can be used in the recombinant DNA construct of the
 CC invention.

XX
 SQ Sequence 664 BP; 157 A; 163 C; 171 G; 173 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 13; Length 664;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAGCTTGATCTT 25
 Db 489 AAAAAAAAAAGCTACCTTGATCTT 465

RESULT 45
 ABQ35845
 ID ABQ35845 standard; DNA; 703 BP.
 XX
 AC ABQ35845;
 XX

DT 12-JUL-2002 (first entry)
 XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22436.
 XX

KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KM drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 KW
 XX

OS Homo sapiens.
 XX

PN WO200218632-A2.
 XX

PD 07-MAR-2002.
 XX

PF 01-SEP-2001; 2001WO-EP010074.
 XX

PR 01-SEP-2000; 2000DE-01043826.
 XX

PR 05-SEP-2000; 2000DE-01044543.
 XX

PA (EPIC-) EPIGENOMICS AG.
 XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX

DR MPI; 2002-371829/40.
 XX

PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX

Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert in a
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridized to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridization to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridized to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNPs); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ1410-
 CC ABQ34121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX

SQ Sequence 703 BP; 290 A; 228 C; 90 G; 95 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 703;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 AAAAAAAAACTATAGCTTGATCTT 25
 |||||
 153 AAAAAAAAACTATAGCTTAATCCT 177

RESULT 46
 ABQ35844/c
 ID ABQ35844 standard; DNA; 703 BP.
 XX
 AC ABQ35844;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22435.
 XX
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KM drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI; 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 PS Claim 12; 56pp + Sequence listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 703 BP; 95 A; 90 C; 228 G; 290 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 6; Length 703;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Query Match 74.4%; Score 18.6; DB 4; Length 774;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 AAAAAAAAACTATAGCTTGATCTT 25
 |||||
 551 AAAAAAAAACTATAGCTTAATCCT 527

RESULT 47
 ABL15043
 ID ABL15043 standard; cDNA; 774 BP.
 XX
 AC ABL15043;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 39611.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 OS
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Ventler JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-6556860/75.
 DR P-PSDB; ABB70940.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 39611; 21pp + Sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB82072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 774 BP; 216 A; 205 C; 203 G; 150 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 4; Length 774;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 AAAAAAAAACTATAGCTTGATCTT 25
 |||||
 588 AAAAAAAAACTATAGCTTGATCTT 612

RESULT 48
 AAC33525/c
 ID AAC33525 standard; DNA; 1157 BP.
 XX
 AC AAC33525;
 XX
 DT 17-OCT-2000 (first entry)
 XX

DE Arabidopsis thaliana DNA fragment SRO ID NO: 3369.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KM promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 16-APR-1999; 99US-0128714P.
PR 19-APR-1999; 99US-0129845P.
PR 21-APR-1999; 99US-0130077P.
PR 23-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135114P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138064P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147703P.
PR 06-AUG-1999; 99US-0147716P.
PR 09-AUG-1999; 99US-0147933P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148865P.
PR 13-AUG-1999; 99US-0148864P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.

```

PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

```

```

Query Match 74.4%; Score 18.6; DB 3; Length 1157;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 25 AAAAAAAAAAGTAGAAGCTTGATTTT 1

```

```

RESULT 49
ABL34328/c
ID ABL34328 standard; DNA; 1160 BP.
XX
XX
AC ABL34328;
XX
XX
DT 26-MAR-2002 (first entry)
XX

```

Human immune system associated gene SEQ ID NO: 2301.

Human; immune system disease; cytosine methylation; antiaesthetic;
 antiarteriosclerotic; antianemic; cytosolic; nootropic;
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

```

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
XX Homo sapiens.
XX WO200200928-A2.
XX
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-BP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIDENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 2301; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/intestine bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 1160 BP; 233 A; 77 C; 363 G; 487 T; 0 U; 0 Other;

```

```

Query Match 74.4%; Score 18.6; DB 6; Length 1160;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAACTATAGCTTGATCTT 25
DB 39 AAAAAAAAACTATAGCTTTAAGCTT 15

```

```

RESULT 50
ABV83585
ID ABV83585 standard; CDNA; 1210 BP.
XX
XX
AC ABV83585;
XX
XX
DT 06-DEC-2002 (first entry)
XX

```

Human breast specific gene SEQ ID NO 28.

Human; breast; cytosolic; cancer; transgenic; gene therapy; vaccine;
 gene; ss.

Homo sapiens.

WO200266605-A2.

29-AUG-2002.

14-FEB-2002; 2002WO-US004284.

15-FEB-2001; 2001US-0268999P.

(DIAD-) DINDXUS INC.

Salceda S, Macina RA, Hu P, Reclipon H, Karra K, Caferkey R;

PI Sun Y, Liu C;
 XX
 DR WPI; 2002-713345/77.
 XX
 PT New isolated breast specific nucleic acid molecules and polypeptides,
 PT useful for identifying, diagnosing, monitoring, staging, imaging and
 PT treating breast cancer and non-cancerous disease states in breast tissue.
 XX
 PS Claim 1, Page 161-162; 254pp; English.
 XX
 CC The invention relates to human breast specific nucleic acids (I)
 CC comprising: (a) a sequence encoding any one of 95 protein sequences
 CC (ABP6614-ABP66708); (b) any one of 115 polynucleotide sequences
 CC (ABV83558-ABV83672); (c) a molecule that selectively hybridizes to (a) or
 CC (b); (d) a molecule having at least 60% sequence identity to (a) or (b).
 CC The breast specific nucleic acid molecules, polypeptides and antibodies
 CC are useful for identifying, diagnosing, monitoring, staging, imaging and
 CC treating breast cancer and non-cancerous disease states in breast tissue.
 CC They are also useful for producing transgenic animals and cells and
 CC producing engineered breast tissue for treatment and research. The
 CC transgenic animals are useful as animal model systems used in elaborating
 CC the biological function of the polypeptides, studying conditions and/or
 CC disorders associated with aberrant expression and in screening for
 CC compounds effective in ameliorating the conditions. The polynucleotides
 CC are useful for gene therapy and in vaccines
 XX
 SQ Sequence 1210 BP; 379 A; 259 C; 288 G; 283 T; 0 U; 1 Other;
 Query Match 74.4%; Score 18.6; DB 6; Length 1210;
 Best Local Similarity 84.0%; Pred. No. 1.2e+03;
 Matches 2; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTATAGCTTGATCTT 25
 Db 437 AAAAAAAAAAGTATAGTTTATATT 461

Search completed: December 14, 2005, 02:42:40
 Job time : 210.2 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:34:03 : Search time 1752.1 Seconds
(without alignments)
667,586 Million cell updates/sec

Title: US-10-681-773-4
Perfect score: 25
Sequence: 1 aaaaaaaaaactatagcttgatctt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.8	87.2	423	9	B2120103 CH230-472
2	21.8	87.2	489	9	BH576123 BOGR18TF
3	21.8	87.2	689	8	CX316160 JGI_X2T66
4	21.8	87.2	756	8	CX316159 JGI_X2T66
5	21.8	87.2	823	8	CX389935 JGI_X2T37
6	21.8	87.2	827	8	CX389936 JGI_X2T37
7	21.8	87.2	1088	11	CNS04Y4Q AL312515 Tetraodon
8	20.8	83.2	515	5	BX557644 BX557644
9	20.8	83.2	593	7	CV137967 ESTR849176
10	20.8	83.2	683	8	DR081344 RTRFPL1_2
11	20.8	83.2	956	8	DR688529 EST107861
12	20.4	81.6	751	10	AG486269 Mus muscu
13	20.4	81.6	877	8	CX406870 JGI_X2T62
14	20.4	81.6	902	10	CZ227059 A1Aa-af3
15	20.2	80.8	202	7	CNS24164 G0015M11
16	20.2	80.8	238	9	B2154934 CH230-243
17	20.2	80.8	282	9	AZ910928 RPT1-24-1
18	20.2	80.8	292	1	AJ676784 AJ676784
19	20.2	80.8	412	8	DN275256 1154331 M
20	20.2	80.8	468	8	BE757975 212569 MA
21	20.2	80.8	487	8	CX244968 1296524 N
22	20.2	80.8	505	6	CB910637 VVD153A06

23	20.2	80.8	505	6	CB918327
24	20.2	80.8	521	3	BJS20168
25	20.2	80.8	531	7	CO879554
26	20.2	80.8	556	3	B1500321
27	20.2	80.8	558	3	B1500177
28	20.2	80.8	570	9	BH733657
29	20.2	80.8	576	9	AZ556893
30	20.2	80.8	583	3	BJS34214
31	20.2	80.8	602	3	BJS32753
32	20.2	80.8	604	3	BJS26127
33	20.2	80.8	637	10	CG43559
34	20.2	80.8	650	9	AZ325230
35	20.2	80.8	659	3	BJS22454
36	20.2	80.8	677	8	CX341975
37	20.2	80.8	686	8	CX408575
38	20.2	80.8	690	3	BJS35981
39	20.2	80.8	703	9	AZ989040
40	20.2	80.8	704	9	BZ469290
41	20.2	80.8	708	10	CE682948
42	20.2	80.8	712	8	CX387836
43	20.2	80.8	715	8	CX579764
44	20.2	80.8	723	8	DNS539361
45	20.2	80.8	731	11	CR896263
46	20.2	80.8	734	8	AZ209019
47	20.2	80.8	740	8	DR898819
48	20.2	80.8	760	10	BX994786
49	20.2	80.8	765	8	CX334205
50	20.2	80.8	766	9	BH564995
51	20.2	80.8	781	11	CR820259
52	20.2	80.8	786	11	CR116777
53	20.2	80.8	791	9	CC484028
54	20.2	80.8	798	10	CZ512816
55	20.2	80.8	809	9	BZ789819
56	20.2	80.8	821	11	CT017777
57	20.2	80.8	825	8	BZ801948
58	20.2	80.8	827	8	CX334206
59	20.2	80.8	833	7	CK773011
60	20.2	80.8	836	8	DN640139
61	20.2	80.8	843	10	BX968009
62	20.2	80.8	850	10	CG002603
63	20.2	80.8	865	10	CZ925732
64	20.2	80.8	865	11	CR094615
65	20.2	80.8	868	10	CZ536632
66	20.2	80.8	872	9	CC683513
67	20.2	80.8	876	10	CG132356
68	20.2	80.8	880	9	CC685385
69	20.2	80.8	885	9	AZ209792
70	20.2	80.8	891	10	AG852293
71	20.2	80.8	891	10	CG119716
72	20.2	80.8	893	5	BX750943
73	20.2	80.8	893	10	CG928442
74	20.2	80.8	900	10	CG500066
75	20.2	80.8	900	10	CG132354
76	20.2	80.8	925	10	CG365056
77	20.2	80.8	932	10	CG500068
78	20.2	80.8	967	6	CF596004
79	20.2	80.8	982	10	CL1994708
80	20.2	80.8	1101	10	CNS00RFG
81	20.2	80.8	1257	10	CL115877
82	20.2	80.8	1257	10	AV163402
83	20.2	80.8	321	10	CNS02ING
84	20.2	80.8	321	10	CM797493
85	19.8	79.2	364	7	CZ266116
86	19.8	79.2	394	1	AV801140
87	19.8	79.2	402	3	BP603671
88	19.8	79.2	402	3	BP603671
89	19.8	79.2	419	3	BP608064
90	19.8	79.2	429	1	AV795931
91	19.8	79.2	433	10	CE778637
92	19.8	79.2	447	3	BP610473
93	19.8	79.2	488	3	DR954497
94	19.8	79.2	535	8	DN245280
95	19.8	79.2	561	11	CT023000

CB918327	VVD031E11
BJS20168	BJS20168
CO879554	BoyGen_07
B1500321	r878e10.y
B1500177	r876f09.y
BH733657	BOWCK56TR
AZ556893	RPT1-23-1
BJS34214	BJS34214
BJS32753	BJS32753
BJS26127	BJS26127
CG43559	OG5EK43TV
AZ325230	1M0047106
BJS22454	BJS22454
CX341975	JGI_X2T46
CX408575	JGI_X2T31
BJS35981	BJS35981
AZ989040	2M0272E16
BZ469290	BONNA66TR
CE682948	Clgr-g88-
CX387836	JGI_X2T21
CX579764	TTE000293
DNS539361	1385130 M
CR896263	Sus scroF
DR898819	JGI_X2T46
AZ209019	SP_0096-B
DN540754	1386666 M
BX994786	Reverse s
CX334205	JGI_X2T69
BH564995	BOGR10TR
CR820259	GROUAA5AE
CR116777	Forward s
CC484028	CH240_313
CZ512816	GMW2-67G1
BZ789819	PURF818TD
CT017777	KBPH129P2
BZ801948	PUGA83TB
CX334206	JGI_X2T69
CK773011	961603 MA
DN640139	UNC-bend
BX968009	Forward s
CG002603	ZUAEE64TV
CZ925732	109842109
CR094615	Forward s
CZ536632	SRAA-aad0
CC683513	OCUDG15TH
CG132356	PUPMY46TD
CC685385	NDL_2P12.
AZ209792	SP_0151_B
AG852293	Oryza bat
CG119716	PUCS01TB
BX750943	BX750943
CG928442	MBECC53TR
CG500066	PUKAS72TB
CG132354	PUPMY46TB
CG365056	MBEHA05TF
CG500068	PUKAS73TD
CF596004	AGENCOURT
CL1994708	ZMHBH1000
AL702030	Droceph11
CL115877	ISBL-6211
AV163402	AV163402
AL197845	Tetraodon
CM797493	WiscDBlox
CZ266116	ME02029.B
AV801140	AV801140
BP603671	BP603671
BP608064	BP608064
AV795931	AV795931
CE778637	Clgr-g88-
BP610473	BP610473
DR954497	ZM_BFD004
DN245280	ACAE-aaas
CT023000	KBPH134K1

C 96	19.8	79.2	644	9	BZ508404	BZ508404	BONAA25TR
C 97	19.8	79.2	647	9	AZ597433	AZ597433	1M0411L10
C 98	19.8	79.2	648	7	CV272564	MS0158.B2	
C 99	19.8	79.2	658	8	CX245248	CX245248	1296839.N
C 100	19.8	79.2	673	7	CV271431	CV271431	W60154.B2
C 101	19.8	79.2	685	5	BY723396	BY723396	BY723396
C 102	19.8	79.2	690	10	AG297491	AG297491	Mus muscu
C 103	19.8	79.2	758	9	BH489771	BH489771	BOCPV707P
C 104	19.8	79.2	758	10	CL770052	CL770052	OR_BBA014
C 105	19.8	79.2	829	10	CG048157	PUGU35TD	
C 106	19.8	79.2	929	10	CNS02VA7	AL115512	Tetradon
C 107	19.8	79.2	944	11	CT013072	KBH12112	
C 108	19.8	79.2	981	9	CC239399	CC239399	CH261-132
C 109	19.8	79.2	983	9	CC355006	CC355006	PURB61TB
C 110	19.8	79.2	1302	10	AUB61104	AUB61104	Brasica
C 111	19.4	77.6	169	5	BUB89310	BUB89310	P019C05.P
C 112	19.4	77.6	341	9	AQ004069	CIT-HSP-2	AA981191
C 113	19.4	77.6	396	1	AA981191	vx58605.r	BY235035
C 114	19.4	77.6	399	5	BY235035	BY235035	HS_5427.A
C 115	19.4	77.6	542	9	AQ683096	AL843980	Dantio.rer
C 116	19.4	77.6	553	11	DR40P11T	CL729848	OR_BBA006
C 117	19.4	77.6	556	10	CL729848	CL187325	CL187325
C 118	19.4	77.6	561	10	CK387947	CK387947	10923B05-
C 119	19.4	77.6	566	7	CK387947	BI378148	BRLG3.001
C 120	19.4	77.6	567	3	BI378148	CW770799	OC_BBA001
C 121	19.4	77.6	576	10	CW770799	CO096695	GR_Ba20B
C 122	19.4	77.6	615	7	CO096695	AG239396	Lotus cor
C 123	19.4	77.6	618	10	AG239396	CW250365	104.712.1
C 124	19.4	77.6	646	10	CW250365	AL704003	DKF2B68T
C 125	19.4	77.6	660	2	BB638136	BB638136	BB638136
C 126	19.4	77.6	660	2	BB638136	CE219989	tigr-g88-
C 127	19.4	77.6	674	9	CE219989	CE219989	tigr-g88-
C 128	19.4	77.6	698	10	CW301499	CW301499	104_785.1
C 129	19.4	77.6	700	10	BX215363	BX215363	Dantio.rer
C 130	19.4	77.6	706	10	CL758438	OR_BBA012	CB946595
C 131	19.4	77.6	755	6	CB946595	AGENCOURT	CW791300
C 132	19.4	77.6	807	10	CW791300	SP_BA007	BU15199
C 133	19.4	77.6	855	5	BU15199	603138951	CC085070
C 134	19.4	77.6	895	9	CC085070	CSU-K33T.	AK017687
C 135	19.4	77.6	1164	4	AK017687	Mus muscu	AK01459
C 136	19.4	77.6	2838	4	AK01459	CL896163	abq36d08.
C 137	19.2	76.8	139	9	CL896163	DN240899	ACAD-aab3
C 138	19.2	76.8	211	8	DN240899	BZ411035	CGABP74TC
C 139	19.2	76.8	238	8	CX066556	BZ411035	CGABP74TC
C 140	19.2	76.8	250	8	CX066556	1321175.N	
C 141	19.2	76.8	251	5	BX614017	CR499737	mtb2-176N
C 142	19.2	76.8	271	11	CR499737	BO456224	ke28a10.Y
C 143	19.2	76.8	334	5	BO456224	CNS60585	tag33a06.
C 144	19.2	76.8	334	7	CNS60585	BP512041	BP512041
C 145	19.2	76.8	352	3	BP512041	AJ605277	AJ605277
C 146	19.2	76.8	369	1	AJ605277	CO280209	EX153408.
C 147	19.2	76.8	375	7	CO280209	DN252005	ACAB-aab6
C 148	19.2	76.8	376	8	DN252005	BE040539	OE08F01.O
C 149	19.2	76.8	378	8	BE040539	BP523768	BP523768
C 150	19.2	76.8	398	3	BP523768		

ALIGNMENTS

RESULT 1
BZ120103
LOCUS BZ120103 423 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-472N3.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
ACCESSION BZ120103
VERSION BZ120103
KEYWORDS BZ120103.1 GI:23761050
SOURCE GSS
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 423)
AUTHORS Zhao, S., Shetty, V., Shatsman, S., Teegaye, G., Geer, K., Chen, D., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Riggs, F., de Jong, P., and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
JOURNAL Other GSSs: CH230-472N3.TJ
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ordering information.html). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 472 row: N column: 3
Seq primer: 17
Class: BAC ends.

FEATURES
source location/Qualifiers
1..423
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SENHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-472N3"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 423;
Best Local Similarity 92.0%; Pred. No. 8.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACATACCTGATCTT 25
366 AAAAAAAAAACTAGCTTCACTT 390

RESULT 2
BH576123/C 489 bp DNA linear GSS 14-DEC-2001
LOCUS BH576123
DEFINITION BOGRC18TF BOGR Brassica oleracea genomic clone BOGRC18, genomic
survey sequence.
ACCESSION BH576123
VERSION BH576123
KEYWORDS BH576123.1 GI:17828404
SOURCE GSS
ORGANISM Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosid II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 489)
AUTHORS Ayele, W., Haas, B.J., Kumar, N., Wu, H., Xiao, Y., Van Aken, S.,
Utebäck, T.R., Wortman, J.R., White, O.R., and Town, C.D.
Whole genome shotgun sequencing of Brassica oleracea and its
application to gene discovery and annotation in Arabidopsis
Genome Res. 15 (4), 487-495 (2005)
JOURNAL 15805480
PUBMED Other GSSs: BOGRC18TR
COMMENT Contact: Chris Town
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208

Email: cdtown@igf.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
Location/Qualifiers

1. 489
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGR18"
/clone_1lb="BOGR"
/note="Vector: pHC01, Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHC01 using BstXI linkers"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 489;
Best Local Similarity 92.0%; Pred. No. 8.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
438 AAAAAAAAACTAGTTGACCTT 414

RESULT 3

CX316160/c
LOCUS CX316160 689 bp mRNA linear EST 04-JAN-2005
DEFINITION JGI_XZT66184.fwd NIH XGC tcriptads Xenopus tropicalis cDNA clone
IMAGE:7785467 5', mRNA sequence.
ACCESSION CX316160
VERSION CX316160.1 GI:57049001

KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.

REFERENCE
AUTHORS 1 (bases 1 to 689)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other_ESTs: JGI_XZT66184.rev
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710

EMAIL: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNLN:
<http://image.llnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0689 row: P column: 9
High quality sequence stop: 626.

FEATURES
source
Location/Qualifiers

1. 689
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAG:7785467"
/tissue_type="whole embryo"

/dev stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_1lb="NIH_XGC tcriptads"
/note="Vector: pCS108, Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
Superscript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector pCS108
(<http://mcb.berkeley.edu/labs/harland/pages/plasmids.html>)

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 689;
Best Local Similarity 92.0%; Pred. No. 8.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
284 AAAAAAAAACTAGCTTGATCTT 260

RESULT 4

CX316159
LOCUS CX316159 756 bp mRNA linear EST 04-JAN-2005
DEFINITION JGI_XZT66184.rev NIH XGC tcriptads Xenopus tropicalis cDNA clone
IMAGE:7785467 3', mRNA sequence.
ACCESSION CX316159
VERSION CX316159.1 GI:57048999

KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.

REFERENCE
AUTHORS 1 (bases 1 to 756)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other_ESTs: JGI_XZT66184.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710

EMAIL: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNLN:
<http://image.llnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0689 row: P column: 9
High quality sequence stop: 756
POLYA=yes.

FEATURES

source

1. 756
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"

/clone="IMAGE:7785467"
/issue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_lib="NIH XGC troptads5"
/note="Vector: PCS108; Site_1: SalI; Site_2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector PCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)
"

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 756;
Best Local Similarity 92.0%; Pred. No. 8.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCTT 25
|||||
472 AAAAAAAAAACATTAGCTGATCTT 496

RESULT 5
LOCUS CX389935 823 bp mRNA linear EST 06-JAN-2005
DEFINITION JGI XZT37948.rev NIH XGC troptads Xenopus tropicalis cDNA clone
ACCESSION CX389935
VERSION CX389935.1 GI:57170613
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 823)
Richardson, P., Lucas, S., Rohsard, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI XZT37948.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cda@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA library preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LNL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Small insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0393 row: h column: 8
High quality sequence scop: 823
POLYATes.

FEATURES
source 1..823
Location/Qualifiers
/organism="Xenopus tropicalis"

/mol_type="mRNA"
/db_xref="taxon:8364"
/IMAGE:7614010"
/clone_lib="NIH XGC troptads5"
/note="Vector: PCS108; Site_1: SalI; Site_2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector PCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)
"

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 823;
Best Local Similarity 92.0%; Pred. No. 8.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCTT 25
|||||
477 AAAAAAAAAACATTAGCTGATCTT 501

RESULT 6
LOCUS CX389936 827 bp mRNA linear EST 06-JAN-2005
DEFINITION JGI XZT37948.fwd NIH XGC troptads Xenopus tropicalis cDNA clone
ACCESSION CX389936
VERSION CX389936.1 GI:57170614
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 827)
Richardson, P., Lucas, S., Rohsard, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI XZT37948.rev
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cda@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA library preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LNL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0393 row: h column: 8
High quality sequence scop: 759.

FEATURES
source 1..827
Location/Qualifiers
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"

```

/clone="IMAGE:761401.0"
/issue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_id="NIH XGC troptad5"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector PCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)
"

ORIGIN
Query Match      87.2%; Score 21.8; DB 8; Length 827;
Best Local Similarity 92.0%; Pred. No. 8.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY
1 AAAAAAAAACTATGCTTGATCTT 25
|||||
349 AAAAAAAAAACATTAGCTTGATCTT 325

RESULT 7
LOCUS      CNS04Y4Q/c      1088 bp      DNA      linear      GSS 26-JUN-2000
DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone
021F09 of library A from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION  ALJ12515.1  GI:9545383
VERSION     ALJ12515.1  GI:9545383
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetradon nigroviridis
ORGANISM   Tetradon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1
AUTHORS    Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fitzames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
TITLE      Estimate of human gene number provided by genome-wide analysis
            using Tetradon nigroviridis DNA sequence
JOURNAL    Nat. Genet. 25 (2), 235-238 (2000)
PUBMED     10835645
REFERENCE   2
AUTHORS    Roest Crolius,H., Jaillon,O., Dasilva,C., Orouf-Costaz,C.,
            Fitzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetradon nigroviridis
JOURNAL    Genome Res. 10 (7), 939-949 (2000)
PUBMED     10899143
REFERENCE   3
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 101 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            Web : www.genoscope.cns.fr)
COMMENT    This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetradon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetradon.
FEATURES
            Location/Qualifiers
                1..1088
                /organism="Tetradon nigroviridis"
                /mol_type="genomic DNA"
                /db_xref="taxon:99883"
                /clone="021F09"
                /clone_1lb="A"

```

```

/note="Genoscope sequence ID : COA021CC05C1
end : T7"

ORIGIN
Query Match      87.2%; Score 21.8; DB 11; Length 1088;
Best Local Similarity 92.0%; Pred. No. 8.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY
1 AAAAAAAAACTATGCTTGATCTT 25
|||||
688 AAAAAAAAACTATGCTTGATCTT 664

RESULT 8
LOCUS      BX557644/c      515 bp      mRNA      linear      EST 10-OCT-2003
DEFINITION BX557644 Glossina morsitans morsitans adult infected gut Glossina
morsitans morsitans cDNA clone Tse33fl1_glc, mRNA sequence.
ACCESSION  BX557644.1  GI:33428825
VERSION     BX557644.1  GI:33428825
KEYWORDS   EST.
SOURCE     Glossina morsitans morsitans
            Glossina morsitans morsitans
ORGANISM   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Hippoboscidae; Glossinidae; Glossina.
REFERENCE   1
AUTHORS    Lehane,M.J., Aksoy,S., Gibson,W., Kethornou,A., Berriman,M.,
            Hamilton,J., Soares,M.B., Bonaldo,M.F., Lehane,S. and Hall,N.
TITLE      Adult midgut expressed sequence tags from the tsetse fly Glossina
            morsitans morsitans and expression analysis of putative immune
            response genes
JOURNAL    Genome Biol. 4 (10), R63 (2003)
PUBMED     14519198
COMMENT    Contact: Hall N
            Pathogen Sequencing Unit
            The Sanger Institute The Wellcome Trust Genome Campus
            Hinxton, Cambridge, CB10 1SA, UK
            Request for clones, please contact: Mike Lehane
            Prof. M.J. Lehane
            School of Biological Sciences,
            University of Wales,
            Bangor LL57 2UW
            All clones with suffix g1c are reverse primer reads starting at 5'
            end of the cDNA all plc reads are from
            the 3' end.
FEATURES
            Location/Qualifiers
                1..515
                /organism="Glossina morsitans morsitans"
                /mol_type="mRNA"
                /sub_species="morsitans"
                /db_xref="taxon:37546"
                /clone="Tse33fl1_glc"
                /issue_type="adult infected gut"
                /clone_1lb="Glossina morsitans morsitans adult infected
                gut"
                /note="country: Zimbabwe; EST from adult gut infected with
                T. brucei"

ORIGIN
Query Match      83.2%; Score 20.8; DB 5; Length 515;
Best Local Similarity 91.7%; Pred. No. 1.9e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY
1 AAAAAAAAACTATGCTTGATCTT 24
|||||
506 AAAAAAAAACTATGCTTGATCTT 483

RESULT 9
LOCUS      CV137967/c      593 bp      mRNA      linear      EST 07-SEP-2004
DEFINITION EST849176 Sequencing ESTs from loblolly pine embryos Pinus taeda

```

ACCESSION CDNA clone RPIAM95 5' end, mRNA sequence.
 VERSION CV137967
 KEYWORDS CV137967.1 GI:51900282
 SOURCE EST.
 ORGANISM Pinus taeda (loblolly pine)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 1 (bases 1 to 593)
 REFERENCE Buell, C.R., Zheng, L., Cowles, A. and Cairney, J.
 TITLE Sequencing of ESTs from loblolly pine embryonic libraries
 JOURNAL Unpublished (2004)
 COMMENT Contact: C. Robin Buell
 Plant Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: rbuell@tigr.org
 This clone is available through TIGR. Please contact pine@tigr.org for further information
 Seq primer: ATT TAG GTG ACA CTA TAG.
 Location/Qualifiers
 1..593
 /organism="Pinus taeda"
 /mol_type="mRNA"
 /cultiivar="7-56 mother tree, open-pollinated tree from, Lyons, Georgia, USA"
 /db_xref="taxon:3352"
 /clone="RPIAM95"
 /lab_host="E.coli DH10B-Tona"
 /note="Organ: Zygotic Embryo and Megagametophyte, Somatic Embryo; Vector: pCMV-SPORT 6.1; Site_1: NotI; Site_2: EcoRV; tissue: Whole megagametophytes isolated from pine seeds, Whole embryos excised from these megagametophytes, whole somatic embryos and suspensor tissue from tissue culture, isolated from cell line A12. Pooled RNA from zygotic embryos, megagametophytes, and somatic embryos was used for library construction. Pine cones were harvested weekly from open-pollinated 7-56 mother trees, collections occurred from 7/01/02 until 10/15/02. Whole megagametophytes were first isolated from pine seeds, and whole embryos excised from these megagametophytes each was flash frozen. Embryo development was assessed using the system of Pullman et al (Pullman GS, Johnson S, Peter G, Cairney J, Xu N. 2003. Loblolly Pine somatic embryogenesis: development of a maturation medium and resulting embryo quality. Plant Cell Reports 21:747-758 (http://link.springer.de/link/service/journals/00299/content/03/00586/). For photographs see Clavatta et al 2001. (Clavatta VT, Morillon R, Pullman GS, Chrispeels M, Cairney J. 2001. An aquaglyceroporin is abundantly expressed early in the development of the suspensor and the embryo proper of loblolly pine (Pinus taeda L.). Plant Physiol. 127: 1556-1567 (http://www.plantphysiol.org/cgi/content/full/127/4/1556))".

ORIGIN
 Query Match 83.2%; Score 20.8; DB 7; Length 593;
 Best Local Similarity 91.7%; Pred. No. 1.9e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTATAGCTTGATCT 24
 |||||
 Db 129 AAAAAAAAACTATAGCTTGCTTCT 106

RESULT 10
 DR081344 683 bp mRNA linear EST 09-JUN-2005
 LOCUS RTFEBPL1_29_B03_b1_A029 Roots plus added iron Pinus taeda cDNA clone
 DEFINITION RTFEBPL1_29_B03_3', mRNA sequence.
 ACCESSION DR081344

VERSION DR081344.1 GI:67195964
 KEYWORDS EST.
 SOURCE Pinus taeda (loblolly pine)
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 1 (bases 1 to 683)
 REFERENCE Pratt, L., Cordomier-Pratt, M.M., Lorenz, W.W., Zimmermann, C. and Dean, J.F.D.
 TITLE An EST database from iron-overloaded loblolly pine (Pinus taeda) roots
 JOURNAL Unpublished (2005)
 COMMENT Other_ESTs: RTFEBPL1_29_B03_g1_A029
 Contact: Cordomier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@uga.edu
 RNA prepared and library constructed by W. Walter Lorenz (School of Forest Resources, University of Georgia); plant material prepared by Craig Zimmermann (School of Forest Resources, University of Georgia) using rooted cuttings provided by the Forest Biology Research Cooperative (FBRC) and the CCLONES project at the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
 Seq primer: M13-21 (GTGAACGACGCCAGT)
 POLYA=No.
 Location/Qualifiers
 1..683
 /organism="Pinus taeda"
 /mol_type="mRNA"
 /strain="3 CCLONES"
 /db_xref="taxon:3352"
 /clone="RTFEBPL1_29_B03_A029"
 /lab_host="DH10B-T1 phage-resistant E. coli"
 /note="Organ: Root; Vector: pSL180; Site_1: EcoRI; Site_2: XhoI; The library was prepared from polyA+ RNA from the roots of 1-year-old loblolly pine (Pinus taeda) cuttings that were rooted and then planted in washed sand. The rooted cuttings were maintained for 143 days (August 1 2003 harvest) under ambient conditions in a local greenhouse. They were kept on a weekly regimen of 0.5x nutrient-complete Hoagland's solution and supplemented with additional water sufficient to maintain a 15% soil moisture content. Twenty-four hours (24h) prior to harvesting roots for mRNA preparation, the potted trees were watered with 500 uM FeSO4 (aq) until the sand was saturated. Double-stranded cDNA was cloned unidirectionally into pSL180. Inserts can be excised with EcoRI (5' end) and XhoI (3' end)."

ORIGIN
 Query Match 83.2%; Score 20.8; DB 8; Length 683;
 Best Local Similarity 91.7%; Pred. No. 1.9e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTATAGCTTGATCT 24
 |||||
 Db 561 AAAAAAAAACTATAGCTTGCTTCT 584

RESULT 11
 DR688529 956 bp mRNA linear EST 12-JUL-2005
 LOCUS EST107613 Normalized pine embryo library, lib_D Pinus taeda cDNA
 DEFINITION clone PMAC742 3' end, mRNA sequence.
 ACCESSION DR688529

VERSION DB688529.1 GI:70777005
 KEYWORDS EST.
 SOURCE Pinus taeda (loblolly pine)
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 REFERENCE 1 (bases 1 to 956)
 AUTHORS Buell, C.R., Zheng, L., Cowles, A. and Cairney, J.
 TITLE Sequencing of ESTs from loblolly pine embryonic libraries
 JOURNAL Unpublished (2004)
 COMMENT Contact: C. Robin Buell
 Plant Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: rbuell@tigr.org
 This clone is available through TIGR. Please contact pine@tigr.org
 for further information
 Seq primer: TAA TAC GAC TCA CTA TAG GG.
 FEATURES
 source
 1..956
 /organism="Pinus taeda"
 /mol_type="mRNA"
 /cultivar="7-56 mother tree, open-pollinated tree from,
 Lyons, Georgia, USA"
 /db_xref="taxon:3352"
 /clone="PWAC742"
 /tissue_type="Whole embryos excised from megagametophytes"
 /lab_host="DH10B"
 /clone_lib="Normalized pine embryo library, lib D"
 /note="Organ: Zygotic Embryos; Vector: pGADT7-RecB;
 Site_1: EcoRI; Site_2: ClaI; library enriched in
 transcripts present in suspension culture of loblolly pine
 (Pinus taeda L.), 7-56. Suspension culture was established
 from fertilized megagametophytes, according to method of
 Pullman et al 2003 (Pullman, G., S. Johnson, G. Peter, J.
 Cairney, and N. Xu. 2003. Improving loblolly pine somatic
 embryo maturation: Comparison of somatic and zygotic
 embryo morphology, germination, and gene expression. Plant
 Cell Reports 21:747-758.) This suspension cell culture,
 which contains early developing embryos and
 undifferentiated embryogenic tissue, was the source of
 the RNA used for the suspension culture library.
 Subtracted from this was cDNA made from a pool of RNA from
 zygotic embryos and female gametophytes from all stages of
 development plus somatic embryos from developmental stages
 beyond suspension culture (see Ciavatta VT, Morrillon R,
 Pullman GS, Christpeels M, Cairney J. 2001. An
 aquaglyceroporin is abundantly expressed early in the
 development of the suspensor and the embryo proper of
 loblolly pine (Pinus taeda L.). Plant Physiol. 127:
 1556-1567
 [http://www.plantphysiol.org/cgi/content/full/127/4/1556])
 ORIGIN
 Query Match 83.2%; Score 20.8; DB 8; Length 956;
 Best Local Similarity 91.7%; Pred. No. 1.9e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAGCTTGATCT 24
 Db 133 AAAAAAAAACTAGCTTGATCT 110
 RESULT 12
 AG486269/c 751 bp DNA linear GSS 22-DEC-2004
 LOCUS AG486269
 DEFINITION Mus musculus molossinus DNA, clone:MSWg01-382F12.T7, genomic survey
 sequence.
 ACCESSION AG486269
 VERSION AG486269.1 GI:48193499
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus (Japanese wild mouse)

ORGANISM Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Abe, K., Noguchi, H., Tagawa, K., Yuzurika, M., Toyoda, A., Kojima, T.,
 Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaiki, K. and
 Shiroishi, T.
 TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to
 genomic constitution of strain C57BL/6J, as defined by BAC-end
 sequence-SNP analysis
 JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
 PUBMED 15574823
 REFERENCE 2 (bases 1 to 751)
 AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 TITLE Direct Submision
 JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
 1-7-22 Suenho-chou, Tsunumi-ku, Yokohama, Kanagawa, 230-0045, Japan
 (E-mail:hattori@gsc.riken.jp, URL:http://hsp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the mouse BAC library MSMG01. For BAC
 library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
 Tsukuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1
 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp
 PRIMERS
 Sequencing : T7
 LIBRARY
 Vector : PBACe3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI.
 FEATURES
 source
 1..751
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMG01-382F12.T7"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSWg01 Mouse Male BAC Library"
 ORIGIN
 Query Match 81.6%; Score 20.4; DB 10; Length 751;
 Best Local Similarity 95.5%; Pred. No. 2.7e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAGCTTGAT 22
 Db 718 AAAAAAAAACTAGCTTGAT 697
 RESULT 13
 CX406870/c 877 bp mRNA linear EST 06-JAN-2005
 LOCUS JGI_XZT62168.fwd NIH_XGC_troptads Xenopus tropicalis cDNA clone
 IMAGE:7636864 5', mRNA sequence.
 ACCESSION CX406870
 VERSION CX406870.1 GI:57187570
 KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 REFERENCE 1 (bases 1 to 877)
 AUTHORS Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
 Brockstein, P. and Lindquist, E.A.
 TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
 JOURNAL Unpublished (2004)

COMMENT

Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicals.berkeley.edu/home>
CDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNLU:
<http://image.lnl.gov>

Naming Conventions: EST name is generated by the concatenation of the UGI Clone Id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
Plate: XZT 0645 row: P column: 14
High quality sequence stop: 794.

FEATURES

source

Location/Qualifiers
1..877
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7636864"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electrogen-Blue"
/clone_lib="NIH XGC troptad5"
/note="vector: PCS108; Site_1: SalI; Site_2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector PCS108
(<http://mcdb.berkeley.edu/labs/harland/pages/plasmids.html>)"

ORIGIN

Query Match 81.6%; Score 20.4; DB 8; Length 877;
Best Local Similarity 95.5%; Pred. No. 2.7e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGAT 22
|||||
384 AAAAAAAAACTATAGCTTGAT 363

Db 384 AAAAAAAAACTATAGCTTGAT 363

RESULT 14 902 bp DNA linear GSS 10-FEB-2005
LOCUS C2227059
DEFINITION A1Aa-af31m04.b1 Ancylostoma caninum whole genome shotgun library (A1AAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION C2227059
VERSION C2227059.1 GI:59254004
KEYWORDS GSS.
SOURCE Ancylostoma caninum (dog hookworm)
ORGANISM Ancylostoma caninum
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida; Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
1 (bases 1 to 902)
Mitreva, M., McCarter, J.P., Pape, D., Ritter, E., Tsagaris, V.I., R., Ronko, I., Martin, J., Wylie, T., Dante, M., Meyer, R., Messina, D., Waterston, R.H., Clifton, S.W. and Wilson, R.
Genome Survey sequences from the parasitic nematode Ancylostoma caninum

REFERENCE 1 (bases 1 to 902)
Mitreva, M., McCarter, J.P., Pape, D., Ritter, E., Tsagaris, V.I., R., Ronko, I., Martin, J., Wylie, T., Dante, M., Meyer, R., Messina, D., Waterston, R.H., Clifton, S.W. and Wilson, R.
Genome Survey sequences from the parasitic nematode Ancylostoma caninum

TITLE Unpublished (2004)
Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 266 1800

JOURNAL
COMMENT

Unpublished (2004)
Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 266 1800

Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA provided by John Hawdon (mtmjhg@wumc.edu) DNA sequenced by Washington University Genome Sequencing Center
Class: shotgun.
Location/Qualifiers
1..902
/organism="Ancylostoma caninum"
/mol_type="genomic DNA"
/strain="Baltimore"
/db_xref="taxon:29170"
/dev_stage="Adult"
/lab_host="GSI0"
/clone_lib="Ancylostoma caninum whole genome shotgun library (A1AAGSS 001)"
/note="vector: pOTW13; Site_1: BstXI; Site_2: BstXI; Ancylostoma caninum genomic DNA was randomly sheared, end-repaired and size fractionated to enrich for 2-4 kb fragments. Genomic DNA was provided by John Hawdon (mtmjhg@wumc.edu) at George Washington University. Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."

FEATURES

source

Query Match 81.6%; Score 20.4; DB 10; Length 902;
Best Local Similarity 95.5%; Pred. No. 2.7e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

QY 1 AAAAAAAAACTATAGCTTGAT 22
|||||
652 AAAAAAAAACTATAGCTTGCT 673

Db 652 AAAAAAAAACTATAGCTTGCT 673

RESULT 15 202 bp mRNA linear EST 28-APR-2004
LOCUS C224164/c
DEFINITION GQ015M11.T3_A06 GQ015 Populus trichocarpa x Populus deltoides cDNA clone GQ015M11_A06 5', mRNA sequence.
ACCESSION C224164
VERSION C224164.1 GI:46842533
KEYWORDS EST.
ORGANISM Populus trichocarpa x Populus deltoides
Populus trichocarpa x Populus deltoides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.
1 (bases 1 to 202)
Morency, M.-J., Cooke, J., Pavy, N., Parsons, L., Paule, C., Seguin, A., Retzel, E. and Mackay, J.
Arborea EST sequencing in Populus sp. (poplar)
Unpublished (2004)
Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Bugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id Identifier: MN5231368 Clone ID:
GQ015M11_A06 Clones available through: John Mackay, Ph. D.
Professeur adjoint -Assistant professeur EMAL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: M11 row: 06 column: A
Seq primer: T3 primer.

FEATURES

source

Location/Qualifiers
1..202
/organism="Populus trichocarpa x Populus deltoides"
/mol_type="mRNA"
/strain="H1-11"
/db_xref="taxon:3695"
/clone="GQ015M11_A06"

ORIGIN

QY 1 AAAAAAAAACTATAGCTTGATCTT 25

```

VERSION      BZ154934.1  GI:237955887

```

Mammalia; Eutheria; Eua

Riggs, F., de Jong, P. an

Department of Eukaryoti

Clones are derived from

source	1. .238
--------	---------

source	1. .238
--------	---------

ORIGIN

Qy

DECEIT T

VERSIC
VERWON

3

ORIGIN

Matc.

QY 1 AAAAAAAAACTATGCTTGATCTT 25
 |||||
 41 AAAAAAAAAAGCAAACTTGATCTT 17

RESULT 18
 AJ676784 292 bp mRNA linear EST 29-JUN-2004
 LOCUS AJ676784 KN224 Bos taurus cDNA clone KN224-022_A14, mRNA sequence.
 DEFINITION AJ676784
 ACCESSION AJ676784.1 GI:49409074
 VERSION EST.
 KEYWORDS Bos taurus (cow)
 SOURCE Bos taurus
 ORGANISM Bos taurus

REFERENCE
 Anderson, S.I., Finlayson, H.A. and Archibald, A.L.
 Development of cDNA and EST resources for studying reproduction and embryo development in pigs and cattle
 Unpublished (2004)
 CONTACT: Anderson SI
 Genomics and Bioinformatics
 Roslin Institute
 Roslin, Midlothian, EH25 9PS, UNITED KINGDOM
 Single pass sequencing. Bases called and trimmed with phred v0.020425.c. Vector identified by cross match with the -minscore 20 and -mismatch 12 options. Vector: pBluescriptII(CK)+ R. Site 1: EcoRI R. Site 2: NotI 5' Seq primer M13P Description: Normalised library constructed from Bovine Uterus tissue. Clones available from UK Centre for Functional Genomics in Farm Animals, Roslin Institute, Roslin, Midlothian, UK, EH25 9PS, www.arkgenomics.org.
 Location/Qualifiers
 1..292
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /clone="KN224-022_A14"
 /issue_type="uterus"
 /clone_lib="KN224"
 /note="Vector: pBluescriptII(CK)+; Site 1: EcoRI; Site 2: NotI; Single pass sequencing. Normalised library constructed from Bovine Uterus tissue."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 1; Length 292;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATGCTTGATCTT 25
 |||||
 61 AAAAAAAAAACTTACTTGATCTT 105

Db

RESULT 19
 DN275256 412 bp mRNA linear EST 02-MAR-2005
 LOCUS DN275256
 DEFINITION 1154331 MARC 7BOV Bos taurus cDNA 3', mRNA sequence.
 ACCESSION DN275256
 VERSION DN275256.1 GI:60443866
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 412)
 Smith, T.P.L., Roberts, A.J., Echtenkamp, S.E., Chitko-McKown, C.G., Wray, J.E. and Keele, J.W.
 A second set of bovine ESTs from pooled-tissue normalized libraries
 Unpublished (2003)
 CONTACT: Smith TPL

TITLE
 JOURNAL
 COMMENT

USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alc option. Vector identified with cross_match v0.990329.
 Plate: RLR8006 row: 1 column: 11
 Seq primer: TAGAAGCACACTGACAG.
 Location/Qualifiers
 1..412
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /issue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 7BOV"
 /note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI; Library made with RNA pooled from multiple tissues including ovary, hindbrain, uterus, and day-30 whole embryos."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 8; Length 412;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATGCTTGATCTT 25
 |||||
 124 AAAAAAAAACTATGCTTGATCTT 148

Db

RESULT 20
 BE757975 468 bp mRNA linear EST 25-APR-2001
 LOCUS BE757975
 DEFINITION 212566 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BE757975
 VERSION BE757975.1 GI:10171967
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 468)
 Smith, T.P.L., Grose, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Cassas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W., Rohrer, G.A., Chitko-McKown, C.G., Pettea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 11282978

TITLE
 JOURNAL
 COMMENT

CONTACT: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCAGTCACGACG
 Plate: 65 row: 1 column: 22
 Seq primer: ATTAGTGACACTATG.
 Location/Qualifiers
 1..468
 /organism="Bos taurus"
 /mol_type="mRNA"

/db xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2BOV"
/note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from testis, thymus,
semilendonsus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 468;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATGCTTGCCTT 25
|||||
456 AAAAAAAAACTATGCTTGCCTT 432

RESULT 21
LOCUS CX244968 487 bp mRNA linear EST 01-MAR-2005
DEFINITION 1296524 NCCCW 02RT Oncorhynchus mykiss cDNA 3', mRNA sequence.
ACCESSION CX244968
VERSION CX244968.1 GI:60361490
KEYWORDS EST.
SOURCE Oncothynchus mykiss (rainbow trout)
ORGANISM Oncothynchus mykiss
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
1 (bases 1 to 487)
Rexroad, C.E., Goupil, A.-S., Guiguen, Y. and Yao, J.
02RT IUS, NCCCW/MWU EST Project, Phase II, in collaboration with
INRA
Unpublished (2004)
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Lleetown Road, Kearneyville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351
Email: crexroad@ncccw.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_aln option. Vector identified with
cross_match v0.990329.
Plate: 100 row: H column: 8
Seq primer: GTATACGACTCCTCATGCG.
Location/Qualifiers
1. 487
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="NCCCW 02RT"
/note="Vector: PCMV Sport6.0; This library was created by
A.-S. Goupil and Y. Guiguen who subcloned the NCCCW 1RT
library from the INRA multi-tissue library."

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 487;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATGCTTGCCTT 25
|||||
25 AAAAAAAAACTATGCTTGCCTT 49

RESULT 22
LOCUS CB910637 505 bp mRNA linear EST 25-APR-2003
DEFINITION VVD153A06_379185 An expressed sequence tag database for abiotic

stressed berries of Vitis vinifera var. Chardonnay Vitis vinifera
cDNA clone VVD153A06 5, mRNA sequence.
CB910637
VERSION CB910637.1 GI:30125298
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; Vitaceae; Vitis.
1 (bases 1 to 505)
Cushman, J.C.
An expressed sequence tag database for abiotic stressed berries of
Vitis vinifera var. Chardonnay
Unpublished (2002)
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer (backward)
Plate: 153 row: A column: 06
Seq primer: T3 20mer
High quality sequence stop: 505.
Location/Qualifiers
1. 505
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"
/clone="VVD153A06"
/tissue_type="berries"
/dev_stage="mixed; 8, 9, 11, 13, 15, 16 weeks dat"
/clone_lib="An expressed sequence tag database for abiotic
stressed berries of Vitis vinifera var. Chardonnay"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI"

ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 505;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATGCTTGCCTT 25
|||||
378 AAAAAAAAACTATGCTTGCCTT 402

RESULT 23
LOCUS CB918327 505 bp mRNA linear EST 25-APR-2003
DEFINITION VVD031E11_347287 An expressed sequence tag database for abiotic
stressed berries of Vitis vinifera var. Chardonnay Vitis vinifera
cDNA clone VVD031E11 5, mRNA sequence.
CB918327
VERSION CB918327.1 GI:30132988
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; Vitaceae; Vitis.
1 (bases 1 to 505)
Cushman, J.C.
An expressed sequence tag database for abiotic stressed berries of
Vitis vinifera var. Chardonnay
Unpublished (2002)
Contact: Cushman JC
Department of Biochemistry
University of Nevada

MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcuohman@unr.edu

PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer (backward)
Plate: 031 row: B column: 11
Seq primer: T3 20mer
High quality sequence stop: 505.
Location/Qualifiers

FEATURES

source

1..505
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"
/clone="VVD031E11"
/issue_type="berries"
/dev_stage="mixed; 8, 9, 11, 13, 15, 16 weeks daf"
/clone_lib="An expressed sequence tag database for abiotic stressed berries of Vitis vinifera var. Chardonnay"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 505;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTGATCTT 25
|||||
Db 378 AAAAAAAAAATTAGCTTCAGCTT 402

RESULT 24 521 bp mRNA linear EST 09-AUG-2002
LOCUS BJ520168/c
DEFINITION BJ520168 MF015SB cDNA Oryzias latipes cDNA clone MF015SB014N19 5',
mRNA sequence.

ACCESSION BJ520168
VERSION BJ520168
KEYWORDS GI:22178380
SOURCE EST.
ORGANISM Oryzias latipes (Japanese medaka)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
JOURNAL Acanthomorpha; Acanthopterygii; Percomorphia; Atherinomorpha;
COMMENT Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 521)

Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES

source

1..521
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF015SB014N19"
/sex="mixture of female and male"
/issue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF015SB cDNA"

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 521;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTGATCTT 25
|||||
Db 236 AAAAAAAAACTGTGATGACTT 212

RESULT 25 551 bp mRNA linear EST 01-SEP-2004
LOCUS CO879554/c
DEFINITION RZPP1056F108Q 5', mRNA sequence.

ACCESSION CO879554
VERSION CO879554
KEYWORDS GI:51809478
SOURCE EST.
ORGANISM Bos taurus (cow)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
JOURNAL Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 551)

Hennig,S., Janitz,M., Herwig,R. and Williams,J.
Generation, annotation, evolutionary analysis and database
integration of 14969 cattle EST clusters
Unpublished (2004)
Contact: Hennig S
Laboraty 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Insestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380

Email: hennig@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones that display the same hybridisation matrix with a
battery of 200 short oligonucleotides are grouped into clusters. One
clone per ONFP cluster was selected for sequencing. cDNA clones and
filters are distributed via Deutsches Ressourcenzentrum fuer
Genomforschung GmbH (<http://www.rzpd.de>).

FORWARD: 5' CCCGAGCTTTACATTATGCTCCGCTCG 3' (M13RSP) 5'-seq
BACKWARD: 5' GCTATTACCGACGCTGGAAGGGCGATG 3' (M13RSP) 3'-seq
Seq primer: 5'-CCGCTCCGCAATTCCTCGGT-3' (M13RSP).

FEATURES

source

1..551
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="RZPP1056F108Q"
/sex="female"
/issue_type="brain tissue"
/dev_stage="adult brain"
/clone_lib="normal cattle brain"
/note="Organ: brain; Vector: pSPORT1; Site 1: NotI;
Site 2: SalI; Random primed and directionally cloned in
pSPORT1 vector using NotI
5'-pGACTAGTTCTAGATCGGACGCGCGCC (T)15-3' and SalI 5'-
TCGACCCACGCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 551;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTGATCTT 25
|||||
Db 416 AAAAAAAAACTATTCATGATCTT 392

RESULT 26 556 bp mRNA linear EST 28-AUG-2001
LOCUS B1500321/c
DEFINITION rs78e10.y1 Prist pacificus mixed stage pamp1 vl Chapelali McCarer

```
ACCESSION      PRistionchus pacificus cDNA 5' similar to SW:TBA2_CAEBL_P34690  
VERSION        BLS00321  
KEYWORDS       TUBULIN ALPHA-2 CHAIN. [1] ; mRNA sequence.  
SOURCE         BLS00321.1 GI:15339665  
ORGANISM       Pristionchus pacificus  
               Neodiplogasteridae; Nematoda; Chromadorea; Diplogasterida;  
REFERENCE      McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,  
AUTHORS          Wyllie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theisling,B.,  
                Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,  
                Tsagarishvili,R., Ronko,I., Kennedy,S., Maguire,U., Beck,C.,  
                Underwood,K., Steptoe,M., Allen,M., Peterson,B., Swaller,T.,  
                Harvey,N., Schurtz,R., Kohn,S., Shih,T., Jackson,Y., Cardenas,M.,  
                McEann,R., Waterston,R. and Wilson,R.  
TITLE           The Washington Univ. Nematode EST Project, 1999  
JOURNAL         Unpublished (1999)  
COMMENT         Contact: McCarter JP  
                 The Washington Univ. Nematode EST project, 1999  
                 Washington University School of Medicine  
                 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
                 Tel.: 314 286 1800  
                 Fax: 314 286 1810  
                 Email: est@watson.wustl.edu  
                 This clone is available royalty-free through LNL ; contact the  
                 IMAGE Consortium (info@image.jnl.gov) for further information.  
                 Trace considered overall poor quality  
FEATURES       High quality sequence stop: 1.  
source         location/Qualifiers  
               1..556  
               /organism="Pristionchus pacificus"  
               /mol_type="mRNA"  
               /db_xref="taxon:54126"  
               /dev_stage="mixed"  
               /lab_host="UDH10B"  
               /clone_lib="Prst pacificus mixed stage pampi vl Chiapellii  
               McCarter"  
               /note="Vector: pAMPl (Gibco) ; Site 1: NotI; Site 2: SalI;  
               The library was constructed by Brandt Chiapelli and Dr.  
               James McCarter at Washington University, St. Louis. The  
               cDNA was made by using Dynabead oligo-dT priming (Dynal).  
               PCR based library using a modified protocol from the  
               SMART PCR cDNA Synthesis Kit from Clontech. Directionally  
               cloned into the UDG sites of pAMPl." 
```

TITLE
 JOURNAL
 COMMENT
 The Washington Univ. Nematode EST Project, 1999
 The Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (lnl@image.llnl.gov) for further information.
 High quality sequence stop: 369.
 Location/Qualifiers
 1..558
 /organism="Pristionchus pacificus"
 /mol_type="mRNA"
 /db_xref="taxon:54126"
 /dev_stage="mixed"
 /lab_host="DHL0B"
 /clone_lib="Prist pacificus mixed stage pAMP1 v1 Chiapelli111"
 /note="Vector: pAMP1 (Gibco) ; Site_1: NotI; Site_2: SalI;
 The library was constructed by Brandi Chapelli and Dr.
 James McCarter at Washington University, St. Louis. The
 PCR was made by using Dynabead oligo-dT priming (Dyna).
 PCR based library using a modified protocol from the
 SMART PCR cDNA Synthesis Kit from Clontech. Directionally
 cloned into the UDG sites of pAMP1. "
 ORIGIN
 Query Match 80.8%; Score 20.2; DB 3; Length 558;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
 QY 1 AAAAAAAAACTATAGCTTGATCTT 25
 |||||
 196 AAAAAAAAAACTCGAGCTTGCTT 172
 RESULT 28
 LOCUS BH733657
 DEFINITION BOMCK56TR BO_2_3_KB Brassica oleracea genomic clone BOMCK56,
 genomic survey sequence.
 ACCESSION BH733657
 VERSION BH733657.1 GI:18839052
 KEYWORDS GSS.
 SOURCE
 ORGANISM
 Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 570)
 Ayele, M., Haas, B.J., Kumar, N., Wu, H., Xiao, Y., Van Aken, S.,
 Utecher, T.R., Wortman, J.R., White, O.R. and Town, C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)
 15805490
 Other GSSs: BOMCK56TF
 CONTACT: Chris Town
 TIGR
 7712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR

Class: sheared ends.
Location/Qualifiers
1..570
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_1lb="BO_2_3_KB"
/note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt using BstXI linkers"

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 570;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCTT 25
|||||
444 AAAAAAAAACTAGCTGATCTT 468

RESULT 29 576 bp DNA linear GSS 20-NOV-2000
AZ556893
LOCUS RPCI-23-179K19.TV RPCI-23 Mus musculus genomic clone
DEFINITION RPCI-23-179K19, genomic survey sequence.
ACCESSION AZ556893
VERSION AZ556893.1 GI:11236713
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 576)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S.,
Akimov, B., Levin, M., Megann, S., Tsengaye, G., Geer, K., Krol, M., de
Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-179K19.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea.ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 179 ROW: K Column: 19
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..576
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-179K19"
/sex="Female"
/lab_host="DH10B"
/clone_1lb="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the

Class: The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies).
ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 576;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCTT 25
|||||
269 AAAAAAAAACTAGCTGATCTT 293

RESULT 30 583 bp mRNA linear EST 09-AUG-2002
BU534214
LOCUS BU534214 MF01SSB CDNA Oryzias latipes cDNA clone MF01SSB019F05 3',
DEFINITION mRNA sequence.
ACCESSION BU534214
VERSION BU534214.1 GI:22193026
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorphae; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adiantichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 583)
Kohara, Y., Shin-I, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES
source
Location/Qualifiers
1..583
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSB019F05"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_1lb="MF01SSB CDNA"

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 583;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCTT 25
|||||
479 AAAAAAAAACTAGCTGATCTT 503

RESULT 31 602 bp mRNA linear EST 09-AUG-2002
BU532753
LOCUS BU532753 MF01SSB CDNA Oryzias latipes cDNA clone MF01SSB014N19 3',
DEFINITION mRNA sequence.
ACCESSION BU532753
VERSION BU532753.1 GI:22191565
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorphae; Acanthopterygii; Percomorpha; Atherinomorpha;

REFERENCE Beloniformes; Adrianchthyidae; Oryziinae; Oryzias.
1 (bases 1 to 602)
AUTHORS Kohara,Y., Shin-1,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehin@genes.nig.ac.jp.

FEATURES
source Location/Qualifiers
1..602
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rs"
/db_xref="taxon:8090"
/clone="MF01SSB014N19"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSB cDNA"

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 602;
Best Local Similarity 88.0%; Pred.No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCTT 25
|||||
Db 499 AAAAAAAAACTATAGCTTGAACTT 523

RESULT 32
BJS26127/c 604 bp mRNA linear EST 09-AUG-2002
LOCUS BJS26127 MF01SSB cDNA Oryzias latipes cDNA MF01SSB019F05 5',
DEFINITION mRNA sequence.
ACCESSION BJS26127
VERSION BJS26127.1 GI:22184939
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Perciformes; Atherinomorpha;
Beloniformes; Adrianchthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 604)
AUTHORS Kohara,Y., Shin-1,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-1
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehin@genes.nig.ac.jp.

FEATURES
source Location/Qualifiers
1..604
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rs"
/db_xref="taxon:8090"
/clone="MF01SSB019F05"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSB cDNA"

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 604;

Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTATAGCTTGATCTT 25
|||||
Db 220 AAAAAAAAACTATAGCTTGAACTT 196

RESULT 33
CG443559 637 bp DNA linear GSS 17-SEP-2003
LOCUS CG443559
DEFINITION CG5EK43TV_ZM_0.7_1.5_KB Zea mays genomic clone ZMMBxa0846h13,
genomic survey sequence.
ACCESSION CG443559
VERSION CG443559.1 GI:34824352
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 637)

REFERENCE Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nuneberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: methylation filtered.
Location/Qualifiers
1..637
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBxa0846h13"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

FEATURES
source

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 637;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCTT 25
|||||
Db 610 AAAAAAAAACTATAGCTTGATCAT 634

RESULT 34
AZ325230 650 bp DNA linear GSS 29-SEP-2000
LOCUS AZ325230
DEFINITION IM0047106R Mouse 10kb plasmid UGCG1M library Mus musculus genomic
clone UGCG1M0047106 R, genomic survey sequence.
ACCESSION AZ325230
VERSION AZ325230.1 GI:10381723
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 650)

REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von

TITLE Niederhausern, A. and Wright, D., Weiss, R.
JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb
COMMENT Plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plates: 0047 row: 1 column: 06
 Seq primer: CACACAGAAACAGCTATGAC
 Class: plasmid ends
 High quality sequence stop: 650.

FEATURES
 source
 1..650
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCIM0047106"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUCIM library"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g1473214|sb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 9; Length 650;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
 |||||
 430 AAAAAAAAAATTATTTGATCTT 454

RESULT 35
 BU522454/c
LOCUS BU522454/c
DEFINITION BU522454 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB024017 5', mRNA sequence.
ACCESSION BU522454
VERSION BU522454.1 GI:22181266
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 659)
 Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.

TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadao Shin-1
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 source
 1..669
 Location/Qualifiers
 /organism="Oryzias latipes"
 /mol_type="mRNA"
 /strain="Hd-IR"
 /db_xref="taxon:8090"
 /clone="MF01SSB024017"
 /sex="mixture of female and male"
 /tissue_type="whole embryo"
 /dev_stage="segmentation stage 20 - 25"
 /clone_lib="MF01SSB cDNA"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 3; Length 669;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
 |||||
 227 AAAAAAAAACTAGTTGAACCT 203

RESULT 36
 CX341975 677 bp mRNA linear EST 04-JAN-2005
 CX341975
LOCUS JGI XZ146425.rev NIH XGC troptred5 Xenopus tropicalis cDNA clone
DEFINITION IMAGE:7621170 3', mRNA sequence.
ACCESSION CX341975
VERSION CX341975.1 GI:57078447
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 677)
 Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C., Brokstein, P. and Lindquist, E.A.
AUTHORS DOE Joint Genome Institute Xenopus tropicalis EST project
TITLE Unpublished (2004)
JOURNAL Contact: Lindquist, E.A., Richardson, P.
COMMENT DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 296 5600
 Fax: 925 296 5710
 Email: cdaa@jgi-psf.org
 Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: http://tropicalis.berkeley.edu/home
 cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
 DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
 Clone Distribution: I.M.A.G.E. Consortium/LNL:
 http://image.llnl.gov
 Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
 Poly-A: Based upon the presence of a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated.
 The resulting Poly-T sequence has been removed.
 Small insert: Based upon one or more sequencing reads of this clone where vector sequence was present at both ends, this clone has been determined to contain a cDNA insert on the order of 600-1000 bases.
 Plate: XZT 0401 row: b column: 16

High quality sequence stop: 658
POLYA=yes

FEATURES

Source

```
Location/Qualifiers
1..677
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7621170"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electrogen-Blue"
/clone_lib="NIH XGC troptAd5"
/note="Vector: pCS108; Site_1: SalI; Site_2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS108 (http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)"
```

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 677;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 472 AAAAAACAAACATTGCTTATCTT 496

RESULT 37
CX408575 686 bp mRNA linear EST 06-JAN-2005
LOCUS JGI XZTJ1688.fwd NIH XGC troptAd5 Xenopus tropicalis cDNA clone
DEFINITION IMAGE:7608028 5', mRNA sequence.

ACCESSION CX408575
VERSION CX408575.1 GI:57189277
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM

REFERENCE
AUTHORS Richardson, P., Lucas, S., Rohrsar, D., Dettler, J.C., Ng, D.C., Brokstein, P. and Lindquist, E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNAL Unpublished (2004)
COMMENT Other ESTs: JGI XZTJ1688.rev
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LNL:
http://image.lnl.gov

Naming Conventions: EST name is generated by the concatenation of the JGI Clone id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
Plate: XZT 0329 row: 0 column: 2
High quality sequence stop: 647.
Location/Qualifiers
1..686
/organism="Xenopus tropicalis"

FEATURES

Source

```
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7608028"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electrogen-Blue"
/clone_lib="NIH XGC troptAd5"
/note="Vector: pCS108; Site_1: SalI; Site_2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS108 (http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)"
```

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 686;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 380 AAAAAACAAACATTGCTTATCTT 356

RESULT 38
BJS35981 690 bp mRNA linear EST 09-AUG-2002
LOCUS BJS35981 MF01SSB cDNA Oryzias latipes cDNA clone MF01SSB024017 3', mRNA sequence.

ACCESSION BJS35981
VERSION BJS35981.1 GI:22194793
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM

REFERENCE
AUTHORS Kohata, Y., Shin-I, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
Source
1..690
Location/Qualifiers
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSB024017"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSB cDNA"

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 690;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 499 AAAAAACAAACATTGCTTATCTT 523

RESULT 39
 AZ989040/c 703 bp DNA linear GSS 27-APR-2001
 LOCUS 2M0272E16F Mouse 10kb plasmid UGCGM library Mus musculus genomic
 DEFINITION clone UGCGM0272E16 F, genomic survey sequence.
 ACCESSION AZ989040
 VERSION AZ989040.1 GI:13860267
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 703)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A. and Wright, D., Weis, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 JOURNAL Contact: Robert B. Weis
 COMMENT University of Utah Genome Center
 Km. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0272 row: B column: 16
 Seq primer: CGTTGTAAACGACGCCACT
 Class: plasmid ends
 High quality sequence stop: 703.
 Location/Qualifiers
 1..703
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UGCGM0272E16"
 /sex="female"
 /lab_host="E. coli strain XL10-Gold, Ti-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UGCGM library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.3 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (g1473214|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptor complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

Query Match 80.8%; Score 20.2; DB 9; Length 703;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 1 AAAAAAAAACTAGCTGATCTT 25
 77 AAAAAAAAAAGAGCTTGATCTT 53

RESULT 40
 B2469290 704 bp DNA linear GSS 13-DEC-2002
 LOCUS BONNA86TR_BO_1.6.2_KB_tot Brassica oleracea genomic clone BONNA86,
 DEFINITION genomic survey sequence.
 ACCESSION B2469290
 VERSION B2469290.1 GI:26765141
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 704)
 Ayele, M., Haas, B.J., Kumar, N., Wu, H., Xiao, Y., Van Aken, S.,
 Utecherback, T.R., Wortman, J.R., White, O.R. and Town, C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)
 15805490
 Other GSSs: BONNA86TF
 JOURNAL Contact: Chris Town
 COMMENT TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
 Location/Qualifiers
 1..704
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BONNA86"
 /clone_lib="BO_1.6.2_KB_tot"
 /note="Vector: PHO81; Site 1; BacXI; 1.6-2 kb sheared
 total DNA inserted into PHO81 using BacXI linkers"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 9; Length 704;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 1 AAAAAAAAACTAGCTGATCTT 25
 576 AAAAAAAAAACCTAGCTGATCTT 600

RESULT 41
 CE682948 708 bp DNA linear GSS 29-SEP-2003
 LOCUS tigr-gss-dog-17000329548762 Dog Library Canis familiaris genomic.
 DEFINITION genomic survey sequence.
 ACCESSION CE682948
 VERSION CE682948.1 GI:37001984
 KEYWORDS GSS.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 Canis.
 1 (bases 1 to 708)
 Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
 Rueda, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
 Venter, J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 14512627

COMMENT

Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org

FEATURES

Class: shotgun.
Location/Qualifiers

ORIGIN

1..708
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_id="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

Query Match 80.8%; Score 20.2; DB 10; Length 708;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 150 AAAAAAAAACTTAGCTTGATGTT 174

RESULT 42

CX387836

LOCUS CX387836 712 bp mRNA linear EST 05-JAN-2005
DEFINITION JGI_XZT21799.rev NIH XGC troptads Xenopus tropicalis cDNA clone
IMAGE:7598404.3, mRNA sequence.

ACCESSION CX387836
VERSION CX387836.1 GI:57156393
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis

REFERENCE Amphibia; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 712)
Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)

OTHER ESTS: JGI_XZT21799.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org

TITLE

JOURNAL

COMMENT

source

Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/BLNI:
<http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of
the JGI clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.

Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.

Plate: XZT 0225 row: n column: 2
High quality sequence stop: 632
POLYA=yes

FEATURES

Location/Qualifiers

1..712
/organism="Xenopus tropicalis"

ORIGIN

/mol_type="mRNA"
/db_xref="taxon:9614"
/clone_id="IMAGE:7598404"
/issue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_id="NIH XGC troptads5"
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dT primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector pCS108
(<http://mc.berkeley.edu/labs/harland/pages/plasmids.html>)
"

Query Match 80.8%; Score 20.2; DB 8; Length 712;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 683 AAAAAAAAAAATAGCTTAATCTT 707

RESULT 43

CX579764

LOCUS CX579764 715 bp mRNA linear EST 13-JAN-2005
DEFINITION TTE00029309 Amplicon Express - Conjugative Form Tetrahymena
thermophila cDNA, mRNA sequence.

ACCESSION CX579764
VERSION CX579764.1 GI:57672295
KEYWORDS EST.
SOURCE Tetrahymena thermophila
ORGANISM Tetrahymena thermophila

REFERENCE Tetrahymena thermophila
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Tetrahymenina; Tetrahymenidae; Tetrahymena.
1 (bases 1 to 715)
Garg, J., Pearlman, R.E. and Carlton, J.
pepdbPub (<http://amoebidia.bcm.umontreal.ca/public/pepdb/agrm.php>)
Tetrahymena thermophila (TIGR)

JOURNAL Unpublished (2004)
COMMENT Contact: pepdb
Departement de Biochimie, Universite de Montreal
Email: pepdb-curatore@ch.umontreal.ca
Plate: 3435.

FEATURES

source

ORIGIN

Location/Qualifiers
1..715
/organism="Tetrahymena thermophila"
/mol_type="mRNA"
/db_xref="taxon:5911"
/clone_id="Amplicon Express - Conjugative Form"

Query Match 80.8%; Score 20.2; DB 8; Length 715;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 225 AAAAAAAAAATAGCTTAATCTT 249

RESULT 44

DN539361/c

LOCUS DN539361 723 bp mRNA linear EST 11-MAR-2005
DEFINITION 1385130 MARC 7BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION DN539361
VERSION DN539361.1 GI:60994156

KEYWORDS EST.
SOURCE Bos taurus (cow)

```

ORGANISM      Bos taurus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
                Pecora; Bovidae; Bovinae; Bos.
TITLE          1 (bases 1 to 723)
JOURNAL        Smith,T.P.L., Roberts,A.J., Echtenkamp,S.E., Chitko-McKown,C.G.,
COMMENT        Wray,J.E. and Keele,J.W.
                A second set of bovine ESTs from pooled-tissue normalized libraries
                Unpublished (2003)
                Contact: Smith TPL
                USDA, ARS, US Meat Animal Research Center
                PO Box 166, Clay Center, NE 68933-0166, USA
                Tel: 402 762 4366
                Fax: 402 762 4390
                Email: smith@mail.marc.usda.gov
                Single pass sequencing. Bases called with phred v0.020425.c and
                trimmed with the aid of the trim_alt option. Vector identified with
                cross_match v0.990329.
                Plate: RUK8068 row: 0 column: 21
                Seq primer: GTAAATACGACTCCTCATAGG.
FEATURES
  source
    1..723
    /organism="Bos taurus"
    /mol_type="mRNA"
    /db_xref="taxon:9913"
    /tissue_type="pooled"
    /lab_host="DH10B"
    /clone_lib="MARC 7BOV"
    /note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
    Library made with RNA pooled from multiple tissues
    including ovary, hindbrain, uterus, and day-30 whole
    embryos."
ORIGIN
Query Match      80.8%; Score 20.2; DB 8; Length 723;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 367 AAAAAAAAACTATTCATGATCTT 343
RESULT 45
CR896263      731 bp      DNA      linear      GSS 22-NOV-2004
LOCUS         CR896263
DEFINITION   Sus scrofa BSS, genomic survey sequence.
ACCESSION    CR896263
VERSION      CR896263.1 GI:55974927
KEYWORDS     GSS; Bac-end sequence BBS; Genome Survey Sequence.
SOURCE       Sus scrofa (pig)
ORGANISM     Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
1 (bases 1 to 731)
Rogel-Galliard,C., Bourgeaux,N., Billault,A., Vaiman,M. and
Chardon,P.
Construction of a swine BAC library: application to the
characterization and mapping of porcine type C endoviral elements
Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)
10449899
2 (bases 1 to 731)
Chardon,P., Iannuccielli,N., Roig,A., Dossat,C., Demars,J.,
Rogel-Galliard,C., Roy,A., Schibler,L. and Milan,D.
A physical map of the swine genome
Unpublished
3 (bases 1 to 731)
Genoscope.
Direct Submission
Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : segreff@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

```

```

FEATURES
  source
    1..731
    /organism="Sus scrofa"
    /mol_type="genomic DNA"
    /strain="Large White"
    /db_xref="taxon:9823"
    /clone_lib="B10227A02"
    /sex="male"
    /cell_type="fibroblast"
    /clone_lib="SBAB"
    /note="Genoscope sequence ID : IH0AAAZ2AC07PM1."
ORIGIN
Query Match      80.8%; Score 20.2; DB 11; Length 731;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAGCTTGATCTT 25
Db 374 AAAAAAAAACTATATGATCTT 398
RESULT 46
DR898819/c      734 bp      mRNA      linear      EST 01-AUG-2005
LOCUS         DR898819/c
DEFINITION   UGI_XZT46425.fwd NIH_XGC_trop1ad5 Xenopus tropicalis cDNA clone
IMAGE:7621170 5', mRNA sequence.
ACCESSION    DR898819
VERSION      DR898819.1 GI:71588071
KEYWORDS     EST.
SOURCE       Xenopus tropicalis (western clawed frog)
ORGANISM     Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 734)
Richardson,P., Lucas,S., Rohrer,D., Dettler,J.C., Ng,D.C.,
Brockstein,P. and Lindquist,B.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: UGI_XZT46425.rev
Contact: Lindquist,B.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley; http://tropicalis.berkeley.edu/home
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute; http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LINL;
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the UGI Clone ID and the direction of sequencing. The suffix '.fwd'
indicates a forward sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0481 row: b column: 16
High quality sequence stop: 707.
location/Qualifiers
    1..734
    /organism="Xenopus tropicalis"
    /mol_type="mRNA"
    /db_xref="taxon:9364"
    /clone_lib="IMAGE:7621170"
    /tissue_type="whole embryo"
    /dev_stage="Tadpole (st. 36-41)"
    /lab_host="E. coli XL1-Blue derivative, Stratagene
    Electrosen-Blue"

```

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT						
DN540754	1386666 MARC 7BOV Bos taurus cDNA 3', mRNA sequence.	DN540754	1	GI:60996934	Bos taurus (cow)	Bos taurus	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	1 (bases 1 to 740)	Smith, T. P. L., Roberts, A. J., Echerkamp, S. E., Chitko-McKown, C. G., Wray, J. E. and Keefe, J. W.	A second set of bovine ESTs from pooled-tissue normalized libraries	Unpublished (2003)						
USDA, ARS, US Meat Animal Research Center	PO Box 166, Clay Center, NE 68933-0166, USA	Tel: 402 762 4366	Fax: 402 762 4390	Email: smtl@mail.marc.usda.gov	Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_al option. Vector identified with cross match v0.990329.	Plate: RLK8068 row: O column: 21	Seq primer: TAGAAGCAGACGTGACG.	Location/Qualifiers	1..740	/organism="Bos taurus"	/mol_type="mRNA"	/db_xref="taxon:9913"	/feature_type="pooled"	/lab_host="DH10B"	/clone_id="MARC 7BOV"	/note="Vector: pCDNA3.1; Site_1: EcoRI; Site_2: NotI; Library made with RNA pooled from multiple tissues including ovary, hindbrain, uterus, and day-30 whole embryos."	
Query Match	80.8%;	Score 20.2;	DB 8;	Length 740;	Best Local Similarity	88.0%;	Pred. No. 3.2e+03;	Matches	22;	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
Qy	1	AAAAAAAACTAGCTGATCTT	25														
Db	374	AAAAAAAAAACTATTCATGATCTT	398														
RESULT 49	EX994786	760 bp	DNA	linear	GSS 05-JUN-2004	Reverse strand read from insert in 3'HPT insertion targeting and chromosome engineering clone MHP405c17, genomic survey sequence.	EX994786	1	GI:49726244	GSS: genome survey sequence; MLCR.	Mus musculus (house mouse)	Mus musculus	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.	1 (bases 1 to 760)	Adams, D. J., Biggs, P. J., Cox, A. V., Davies, R. M., van der Weyden, L., Jonkers, J., Smith, J., Plumb, R. W., Taylor, R. G., Nishijima, T., Yu, Y., Rogers, J. and Bradley, A.	Direct Submision	Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. http://www.sanger.ac.uk/MICR

FEATURES

Location/Qualifiers
1..760
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHP405C17"
/clone_1b="MHP"

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 760;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 592 AAAAAAAAACTAGCTTGATATT 616

RESULT 50

CX334205 765 bp mRNA linear EST 04-JAN-2005
LOCUS JGI_XZT69768.rev NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone
DEFINITION IMAGE:7788908 3', mRNA sequence.
ACCESSION CX334205
VERSION CX334205.1 GI:57070677
KEYWORDS EST.

SOURCE

Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 765)

Richardson, P., Lucas, S., Roksar, D., Deter, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI_XZT69768.fwd

Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
CDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNL:

<http://image.jnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Place: XZT 0725 row: 0 column: 18
High quality sequence stop: 703
POLYA=Yes.

FEATURES

Location/Qualifiers
1..765
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7788908"
/issue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_1b="NIH_XGC_tropTad5"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 765;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAGCTTGATCTT 25
|||||
Db 676 AAAAAAAAAAAATAGCTTAATCTT 700

Search completed: December 14, 2005, 07:34:55
Job time : 1760.1 secs

lab using poly A RNA and oligo dT primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector PCS108
(<http://mcb.berkeley.edu/labs/harland/pages/plasmids.html>)


```
C 97 17 68.0 3288 6 US-10-750-185-35792 Sequence 35792, A
C 98 17 68.0 3509 9 US-11-077-550-19 Sequence 19, Appl
C 99 17 68.0 3820 6 US-10-750-185-64002 Sequence 64002, A
C 100 17 68.0 4399 6 US-10-750-185-52104 Sequence 52104, A
C 101 17 68.0 5371 6 US-10-821-234-274 Sequence 274, Appl
C 102 17 68.0 5390 6 US-10-849-438-4 Sequence 4, Appl
C 103 17 68.0 13682 6 US-10-995-561-13471 Sequence 13471, A
C 104 17 68.0 18540 6 US-10-995-561-13323 Sequence 13323, A
C 105 17 68.0 20773 6 US-10-995-561-13322 Sequence 13322, A
C 106 17 68.0 21332 6 US-10-995-561-13418 Sequence 13418, A
C 107 17 68.0 33042 6 US-10-995-561-13349 Sequence 13349, A
C 108 17 68.0 33042 6 US-10-995-561-13340 Sequence 13340, A
C 109 17 68.0 52520 6 US-10-995-561-13340 Sequence 13340, A
C 110 17 68.0 68123 6 US-10-995-561-13348 Sequence 13348, A
C 111 17 68.0 110000 6 US-11-155-492-11 Sequence 11, Appl
C 112 17 68.0 135019 6 US-11-112-908-26 Sequence 26, Appl
C 113 17 68.0 150173 7 US-11-112-908-44 Sequence 44, Appl
C 114 17 68.0 167116 7 US-11-112-908-27 Sequence 27, Appl
C 115 17 68.0 171247 7 US-11-112-908-25 Sequence 25, Appl
C 116 17 68.0 172781 7 US-11-112-908-25 Sequence 25, Appl
C 117 17 68.0 173995 6 US-11-121-086-45 Sequence 45, Appl
C 118 17 68.0 182303 6 US-11-121-086-45 Sequence 45, Appl
C 119 17 68.0 191350 6 US-11-112-908-32 Sequence 32, Appl
C 120 17 68.0 193363 7 US-11-112-908-34 Sequence 34, Appl
C 121 17 68.0 197781 7 US-11-112-908-33 Sequence 33, Appl
C 122 17 68.0 217623 6 US-10-995-561-13327 Sequence 23, Appl
C 123 17 68.0 244196 6 US-10-995-561-13327 Sequence 23, Appl
C 124 17 68.0 260209 6 US-10-995-561-13327 Sequence 23, Appl
C 125 17 68.0 305312 6 US-10-995-561-13327 Sequence 23, Appl
C 126 17 68.0 340000 7 US-11-102-978-3 Sequence 3, Appl
C 127 17 68.0 387780 6 US-10-995-561-13259 Sequence 3, Appl
C 128 17 68.0 611587 7 US-11-117-187-209 Sequence 209, Appl
C 129 17 68.0 611587 7 US-11-117-187-209 Sequence 209, Appl
C 130 16.8 67.2 673 6 US-10-750-185-42493 Sequence 42493, A
C 131 16.8 67.2 673 6 US-10-750-185-27322 Sequence 27322, A
C 132 16.8 67.2 139054 7 US-11-121-086-96 Sequence 96, Appl
C 133 16.8 67.2 645179 6 US-10-995-561-13293 Sequence 13293, A
C 134 16.8 67.2 645179 6 US-10-995-561-13293 Sequence 13293, A
C 135 16.6 66.4 201 6 US-10-995-561-52967 Sequence 52967, A
C 136 16.6 66.4 201 6 US-10-995-561-76911 Sequence 76911, A
C 137 16.6 66.4 201 6 US-10-995-561-76911 Sequence 76911, A
C 138 16.6 66.4 201 6 US-10-995-561-77595 Sequence 77595, A
C 139 16.6 66.4 201 6 US-10-995-561-77595 Sequence 77595, A
C 140 16.6 66.4 201 6 US-10-995-561-77595 Sequence 77595, A
C 141 16.6 66.4 600 6 US-10-750-185-4514 Sequence 4514, Ap
C 142 16.6 66.4 600 6 US-10-750-185-4514 Sequence 4514, Ap
C 143 16.6 66.4 650 6 US-10-750-185-27919 Sequence 27919, A
C 144 16.6 66.4 743 6 US-10-750-185-37774 Sequence 37774, A
C 145 16.6 66.4 745 6 US-10-750-185-47284 Sequence 47284, A
C 146 16.6 66.4 963 6 US-10-957-569-7 Sequence 7, Appl
C 147 16.6 66.4 1009 6 US-10-750-185-33776 Sequence 33776, A
C 148 16.6 66.4 1078 6 US-10-750-185-47813 Sequence 47813, A
C 149 16.6 66.4 1103 6 US-10-750-185-57665 Sequence 57665, A
C 150 16.6 66.4 1282 6 US-10-750-185-38004 Sequence 38004, A
```

ALIGNMENTS

```
RESULT 1
US-11-112-908-23
; Sequence 23, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
```

```
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: Patent version 3.3
; SEQ ID NO 23
; LENGTH: 188682
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-23

Query Match 79.2%; Score 19.8; DB 7; Length 188682;
Best Local Similarity 91.3%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATGCTTGATC 23
DB 124820 AAAAAAAAACTTATCTTGATC 124842

RESULT 2
US-10-750-185-47179
; Sequence 47179, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patent version 3.1
; SEQ ID NO 47179
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-47179

Query Match 77.6%; Score 19.4; DB 6; Length 1599;
Best Local Similarity 95.2%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATGCTTGGA 21
DB 1006 AAAAAAAAACAATGCTTGA 1026

RESULT 3
US-11-102-978-3/C
; Sequence 3, Application US/11102978
; Publication No. US20050250142A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Technology Transfer Office
; APPLICANT: University of Utah Research Foundation
; TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
; FILE REFERENCE: 0274-5537, US
; CURRENT APPLICATION NUMBER: US/11/102,978
; PRIOR FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: PCT/US2003/033152
; PRIOR FILING DATE: 2003-10-18
; PRIOR APPLICATION NUMBER: 60/419,576
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 13
```


US-11-121-086-98

Query Match 74.4%; Score 18.6; DB 7; Length 171732;
 Best Local Similarity 84.0%; Pred. No. 2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
 |||||

DB 120361 AAAAAAAAAATTAAGCTTGATCTT 120385

US-11-121-086-16

RESULT 8
 ; Sequence 16, Application US/11121086
 ; Publication No. US20050266459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: POULSEN, TIM S.
 ; APPLICANT: NIELSEN, KIRSTEN V.
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
 ; FILE REFERENCE: 09138.6000-00000
 ; CURRENT FILING DATE: 2005-05-04
 ; PRIOR APPLICATION NUMBER: 60/567,570
 ; PRIOR FILING DATE: 2004-05-04
 ; NUMBER OF SEQ ID NOS: 107
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 16
 ; LENGTH: 189539
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
 US-11-121-086-16

Query Match 74.4%; Score 18.6; DB 7; Length 189539;
 Best Local Similarity 84.0%; Pred. No. 2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
 |||||

DB 188341 AAAAAAAAAATGAATCTTGATCTT 188365

US-10-933-025-22

RESULT 9
 ; Sequence 22, Application US/10933025
 ; Publication No. US20050265987A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ROSEN, STEVEN
 ; APPLICANT: HEMMERICH, STEFAN
 ; APPLICANT: TOMITA, MEGUMI
 ; TITLE OF INVENTION: Sulfotransferases and methods of use
 ; FILE REFERENCE: UCAL-230CON
 ; CURRENT APPLICATION NUMBER: US/10/933,025
 ; CURRENT FILING DATE: 2004-09-01
 ; PRIOR APPLICATION NUMBER: 10/025,966
 ; PRIOR FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 60/258,577
 ; PRIOR FILING DATE: 2000-12-27
 ; PRIOR APPLICATION NUMBER: 60/267,831
 ; PRIOR FILING DATE: 2001-09-02
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 22
 ; LENGTH: 268685
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(268685)
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-933-025-22

Query Match 74.4%; Score 18.6; DB 6; Length 268685;

Best Local Similarity 84.0%; Pred. No. 2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
 |||||

DB 131323 AAAAAAAAAATTAAGCTTGATCTT 131347

US-10-995-561-13396

RESULT 10
 ; Sequence 13396, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF
 ; FILE REFERENCE: CI001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13396
 ; LENGTH: 398287
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(398287)
 ; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
 US-10-995-561-13396

Query Match 74.4%; Score 18.6; DB 6; Length 398287;
 Best Local Similarity 84.0%; Pred. No. 2.1e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATCTT 25
 |||||

DB 249105 AATAAGAAACTAATAGCTTGATTTT 249129

US-10-750-185-57463/C

RESULT 11
 ; Sequence 57463, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: PANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM11100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 57463
 ; LENGTH: 1592
 ; TYPE: DNA
 ; ORGANISM: Bovine
 ; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
 US-10-750-185-57463

Query Match 72.8%; Score 18.2; DB 6; Length 1592;
 Best Local Similarity 87.0%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTTGATC 23
 |||||

Db 100 AAAAAAAAACTATAGATTAC 78

RESULT 12

US-10-750-185-40315
; Sequence 40315, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40315
; LENGTH: 1433
; TYPE: DNA
; ORGANISM: Bovine 19866881121145
US-10-750-185-40315

Query Match 71.2%; Score 17.8; DB 6; Length 1433;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTGA 21
Db 183 AAAAAAAAACTTAAGCTGA 203

RESULT 13

US-10-750-185-60997
; Sequence 60997, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60997
; LENGTH: 1689
; TYPE: DNA
; ORGANISM: Bovine 19866880638112
US-10-750-185-60997

Query Match 71.2%; Score 17.8; DB 6; Length 1689;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AAAAACTATAGCTGATCTT 25
Db 1388 AAAACACTGAGCTGATCTT 1408

RESULT 14

US-10-995-561-13330/c
; Sequence 13330, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13330
; LENGTH: 101046
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(101046)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13330

Query Match 71.2%; Score 17.8; DB 6; Length 101046;
Best Local Similarity 90.5%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTGA 21
Db 53220 AAAAAAAAACTGAGCTGA 53200

RESULT 15

US-11-121-086-98/c
; Sequence 98, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 98
; LENGTH: 171732
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-98

Query Match 71.2%; Score 17.8; DB 7; Length 171732;
Best Local Similarity 90.5%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTGA 21
Db 67264 AAAAAAAAACTATAGCTGA 67244

RESULT 16

US-11-121-086-82/c
; Sequence 82, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086

```
/ CURRENT FILING DATE: 2005-05-04
/ PRIOR APPLICATION NUMBER: 60/567,570
/ PRIOR FILING DATE: 2004-05-04
/ NUMBER OF SEQ ID NOS: 107
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO: 82
/ LENGTH: 193084
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-121-086-82

Query Match          71.2%; Score 17.6; DB 6; Length 193084;
Best Local Similarity 90.5%; Pred. No. 3.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCTGA 21
   |||||
Db 128148 AAAAAAAAACTAGCTGATCTGA 128128

RESULT 17
US-10-750-185-11947
/ Sequence 11947, Application US/10750185
/ Publication No. US2005026063A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 11947
/ LENGTH: 28
/ TYPE: DNA
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Reverse Primer
US-10-750-185-11947

Query Match          70.4%; Score 17.6; DB 6; Length 28;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCT 24
   |||||
Db 3 AAAAAAAAACTAGCTGATCTAT 26

RESULT 18
US-10-995-561-28493
/ Sequence 28493, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CI001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 28493
/ LENGTH: 201
```

```
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-28493

Query Match          70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCT 24
   |||||
Db 94 AAAAAAAAACTAGCTGATCT 117

RESULT 19
US-10-995-561-31568
/ Sequence 31568, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CI001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 31568
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-31568

Query Match          70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAACTAGCTGATCT 24
   |||||
Db 30 AAAAAAAAACTAGCTGATCT 53

RESULT 20
US-10-750-185-2211/C
/ Sequence 2211, Application US/10750185
/ Publication No. US2005026063A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 2211
/ LENGTH: 600
/ TYPE: DNA
/ ORGANISM: Bovine
US-10-750-185-2211

Query Match          70.4%; Score 17.6; DB 6; Length 600;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAGCTGATCT 24
```

Db 348 AAAAAGAACTATAGCATCATAT 325

```

RESULT 21
US-10-750-185-62039
; Sequence 62039, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62039
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Bovine 19866881559877
US-10-750-185-62039

```

Query Match 70.4%; Score 17.6; DB 6; Length 678;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAGAACTATAGCATCATCT 24
 Db 605 AAAAAGAACTATAGCATCATCT 628

```

RESULT 22
US-10-750-185-46337/c
; Sequence 46337, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46337
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Bovine 19866881209625
US-10-750-185-46337

```

Query Match 70.4%; Score 17.6; DB 6; Length 798;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAGAACTATAGCATCATCT 24
 Db 502 AAAAAGAACTATAGCATCATAT 479

```

RESULT 23
US-10-750-185-49314/c
; Sequence 49314, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49314
; LENGTH: 817
; TYPE: DNA
; ORGANISM: Bovine 19866881113106
US-10-750-185-49314

```

Query Match 70.4%; Score 17.6; DB 6; Length 817;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAGAACTATAGCATCATCT 24
 Db 350 AAAAAGAACTATAGCATCATCT 327

```

RESULT 24
US-10-750-185-54504/c
; Sequence 54504, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54504
; LENGTH: 905
; TYPE: DNA
; ORGANISM: Bovine 19866881526848
US-10-750-185-54504

```

Query Match 70.4%; Score 17.6; DB 6; Length 905;
 Best Local Similarity 83.3%; Pred. No. 2.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAGAACTATAGCATCATCT 24
 Db 586 AAAAAGAACTATAGCATCATCT 563

RESULT 25
 US-10-750-185-46768

```
/ Sequence 46768, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ PRIOR FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 46768
/ LENGTH: 1145
/ TYPE: DNA
/ ORGANISM: Bovine 19866881032789
US-10-750-185-46768

Query Match          70.4%; Score 17.6; DB 6; Length 1145;
Best Local Similarity 83.3%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 288 AAAAAAAAAATCTACAGCTTGATTT 311

RESULT 26
US-10-750-185-37306
/ Sequence 37306, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ PRIOR FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 37306
/ LENGTH: 1252
/ TYPE: DNA
/ ORGANISM: Bovine 19866880777097
US-10-750-185-37306

Query Match          70.4%; Score 17.6; DB 6; Length 1252;
Best Local Similarity 83.3%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 222 AAAAAAAAAAGTATAGTATGATGT 245

RESULT 27
US-10-750-185-41839
/ Sequence 41839, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ PRIOR FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 41839
/ LENGTH: 1307
/ TYPE: DNA
/ ORGANISM: Bovine 19866881129411
US-10-750-185-41839

Query Match          70.4%; Score 17.6; DB 6; Length 1307;
Best Local Similarity 83.3%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 370 AAAAAAAAAATCAATGCTTGCTCT 393

RESULT 28
US-10-750-185-60870
/ Sequence 60870, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ PRIOR FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 60870
/ LENGTH: 1307
/ TYPE: DNA
/ ORGANISM: Bovine 19866881174648
US-10-750-185-60870

Query Match          70.4%; Score 17.6; DB 6; Length 1307;
Best Local Similarity 83.3%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTATAGCTTGATCT 24
Db 737 AAAAAAAAAAGTATATTATTTATCT 760

RESULT 29
US-10-750-185-45778
/ Sequence 45778, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
```

```

; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45778
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Bovine 1986688133643
US-10-750-185-45778

Query Match          70.4%; Score 17.6; DB 6; Length 1366;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 155 AAAAAAAAACTATGAGTGATCT 178

RESULT 30
US-10-750-185-27123
; Sequence 27123, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27123
; LENGTH: 1928
; TYPE: DNA
; ORGANISM: Bovine 19866880627663
US-10-750-185-27123

Query Match          70.4%; Score 17.6; DB 6; Length 1928;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 952 AAAAAAAAAACATGTGCTAGATCT 975

RESULT 31
US-10-750-185-40402
; Sequence 40402, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
```

```

; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40402
; LENGTH: 2030
; TYPE: DNA
; ORGANISM: Bovine 19866880791331
US-10-750-185-40402

Query Match          70.4%; Score 17.6; DB 6; Length 2030;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 1345 AAAAAAAAACTGATCTTAATCT 1368

RESULT 32
US-10-750-185-34042/c
; Sequence 34042, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34042
; LENGTH: 2107
; TYPE: DNA
; ORGANISM: Bovine 19866880197959
US-10-750-185-34042

Query Match          70.4%; Score 17.6; DB 6; Length 2107;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 1668 AAAAAAAAAATATATCTTATAT 1645

RESULT 33
US-10-750-185-48677
; Sequence 48677, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
```

```
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 48677
/ LENGTH: 3789
/ TYPE: DNA
/ ORGANISM: Bovine 1986680793656
US-10-750-185-48677

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 6; Length 3789;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 433 AAAAAAAAACTATAGCTTGATCT 456

RESULT 34
US-10-995-561-13256
/ Sequence 13256, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13256
/ LENGTH: 55826
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-13256

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 6; Length 55826;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 4896 AAAAAAAAACTATAGCTTGATCT 4919

RESULT 35
US-11-112-908-35
/ Sequence 35, Application US/1112908
/ Publication No. US2005026659A1
/ GENERAL INFORMATION:
/ APPLICANT: Harris, Cole
/ APPLICANT: Davis, Lisa M.
/ TITLE OF INVENTION: Breast Cancer Biomarkers
/ FILE REFERENCE: 04-164-US
/ CURRENT APPLICATION NUMBER: US/11/112,908
/ CURRENT FILING DATE: 2005-04-22
/ PRIOR APPLICATION NUMBER: US 60/564,758
/ PRIOR FILING DATE: 2004-04-23
/ PRIOR APPLICATION NUMBER: US 60/575,978
/ PRIOR FILING DATE: 2004-06-01
/ PRIOR APPLICATION NUMBER: US 60/631,702
/ PRIOR FILING DATE: 2004-11-30
/ PRIOR APPLICATION NUMBER: US 60/633,826
/ PRIOR FILING DATE: 2004-12-07
/ NUMBER OF SEQ ID NOS: 511
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 35
/ LENGTH: 127340
```

```
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-112-908-35

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 7; Length 127340;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTATAGCTTGATCTT 25
Db 114741 AAAAAAAAACTATAGCTTGATCTT 114764

RESULT 36
US-11-121-086-106
/ Sequence 106, Application US/11121086
/ Publication No. US20050266459A1
/ GENERAL INFORMATION:
/ APPLICANT: POULSEN, TIM S.
/ APPLICANT: NIELSEN, KIRSTEN V.
/ TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
/ FILE REFERENCE: 09138.6000-00000
/ CURRENT APPLICATION NUMBER: US/11/121,086
/ CURRENT FILING DATE: 2005-05-04
/ PRIOR APPLICATION NUMBER: 60/567,570
/ PRIOR FILING DATE: 2004-05-04
/ NUMBER OF SEQ ID NOS: 107
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 106
/ LENGTH: 179777
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-121-086-106

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 7; Length 179777;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTATAGCTTGATCT 24
Db 125070 AAAAAAAAACTATAGCTTGATCT 125093

RESULT 37
US-11-121-086-31
/ Sequence 31, Application US/11121086
/ Publication No. US20050266459A1
/ GENERAL INFORMATION:
/ APPLICANT: POULSEN, TIM S.
/ APPLICANT: NIELSEN, KIRSTEN V.
/ TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
/ FILE REFERENCE: 09138.6000-00000
/ CURRENT APPLICATION NUMBER: US/11/121,086
/ CURRENT FILING DATE: 2005-05-04
/ PRIOR APPLICATION NUMBER: 60/567,570
/ PRIOR FILING DATE: 2004-05-04
/ NUMBER OF SEQ ID NOS: 107
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 31
/ LENGTH: 218821
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (106949)..(106949)
/ OTHER INFORMATION: a, c, g, t, unknown or other
/ NAME/KEY: modified_base
/ LOCATION: (110322)..(110324)
/ OTHER INFORMATION: a, c, g, t, unknown or other
/ NAME/KEY: modified_base
/ LOCATION: (115133)..(115133)
```

OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131300)..(131300)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (139059)..(139158)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157740)..(157740)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157777)..(157777)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157900)..(157900)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157919)..(157919)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (157926)..(157926)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158094)..(158094)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158138)..(158138)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158193)..(158193)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158241)..(158242)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158259)..(158259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158278)..(158278)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158295)..(158295)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158740)..(158839)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158929)..(158929)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163528)..(163530)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163550)..(163550)
OTHER INFORMATION: a, c, g, t, unknown or other

FEATURE:
NAME/KEY: modified_base
LOCATION: (163765)..(163765)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164000)..(164000)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164047)..(164047)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164084)..(164084)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167233)..(167233)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167236)..(167236)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167238)..(167238)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170247)..(170247)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170249)..(170250)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170252)..(170253)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170259)..(170259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170262)..(170263)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170266)..(170266)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174470)..(174470)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174472)..(174472)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174474)..(174474)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179059)..(179060)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179064)..(179064)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:

```
/ NAME/KEY: modified_base
/ LOCATION: (197001)..(197001)
/ OTHER INFORMATION: a, c, g, t, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (197004)..(197005)
/ OTHER INFORMATION: a, c, g, t, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (197007)..(197007)
/ OTHER INFORMATION: a, c, g, t, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (200349)..(200349)
/ OTHER INFORMATION: a, c, g, t, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (200351)..(200351)
/ OTHER INFORMATION: a, c, g, t, unknown or other
/ FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (212425)..(212426)
/ OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-31
```

```
Query Match          70.4%; Score 17.6; DB 7; Length 218821;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTATAGCTGATCT 24
Db      157186 AAAAAAATAAAGCTTGATTT 157209
```

```
RESULT 38
US-10-995-561-13244
/ Sequence 13244, Application US/10995561
/ Publication No. US2005022054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ TITLE OF INVENTION: DETECTION AND USES THEREOF
/ FILE REFERENCE: CI001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13244
/ LENGTH: 222094
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-13244
```

```
Query Match          70.4%; Score 17.6; DB 6; Length 222094;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTATAGCTGATCT 24
Db      90938 AAAAAAAGTAAGTGTGATCT 90961
```

```
RESULT 39
US-10-750-185-59663/c
/ Sequence 59663, Application US/10750185
/ Publication No. US2005026060A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
```

```
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 59663
/ LENGTH: 1130
/ TYPE: DNA
/ ORGANISM: Bovine 19866880629549
US-10-750-185-59663
```

```
Query Match          69.6%; Score 17.4; DB 6; Length 1130;
Best Local Similarity 94.7%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTATAGCTT 19
Db      1095 AAAAAAAGCTATAGCTT 1077
```

```
RESULT 40
US-10-750-185-42555/c
/ Sequence 42555, Application US/10750185
/ Publication No. US2005026060A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 42555
/ LENGTH: 1645
/ TYPE: DNA
/ ORGANISM: Bovine 19866880675159
US-10-750-185-42555
```

```
Query Match          69.6%; Score 17.4; DB 6; Length 1645;
Best Local Similarity 94.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 AAAAAAAAACTATAGCTT 19
Db      1552 AAAAAAAGTAAGCTT 1534
```

```
RESULT 41
US-10-750-185-38854/c
/ Sequence 38854, Application US/10750185
/ Publication No. US2005026060A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
```



```
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38854
; LENGTH: 1720
; TYPE: DNA
; ORGANISM: Bovine 19866881289184
US-10-750-185-38854

Query Match          69.6%; Score 17.4; DB 6; Length 1720;
Best Local Similarity 94.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGTATAGCTT 19
Db 90 AAAAAAAAAAAGTATAGCTT 72

RESULT 42
US-11-121-086-95/c
; Sequence 95, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 95
; LENGTH: 212716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-95

Query Match          69.6%; Score 17.4; DB 7; Length 212716;
Best Local Similarity 94.7%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGTATAGCTT 19
Db 160169 AAAAAAAAAAAGTATAGCTT 160151

RESULT 43
US-10-995-561-18622/c
; Sequence 18622, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18622
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-18622

Query Match          68.8%; Score 17.2; DB 6; Length 201;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Best Local Similarity 79.2%; Pred. No. 2.9e+02;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGTATAGCTTATCT 24
Db 113 AAAAAAAAAAAGTATAGCTTATCTT 90

RESULT 44
US-10-995-561-50175/c
; Sequence 50175, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50175
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-50175

Query Match          68.8%; Score 17.2; DB 6; Length 201;
Best Local Similarity 86.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGTATAGCTTGAT 22
Db 94 AAAAAAAAAAAGTATAGCTTGAT 73

RESULT 45
US-10-995-561-68643/c
; Sequence 68643, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68643
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-68643

Query Match          68.8%; Score 17.2; DB 6; Length 201;
Best Local Similarity 86.4%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGTATAGCTTGAT 22
Db 92 AAAAAAAAAAAGTATAGCTTGAT 71

RESULT 46
US-10-750-185-50331/c
; Sequence 50331, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
```

```

; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50331
; LENGTH: 991
; TYPE: DNA
; ORGANISM: Bovine 19866881539364
US-10-750-185-50331
```

```
Query Match      68.8%; Score 17.2; DB 6; Length 991;
Best Local Similarity 86.4%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTATAGCTTGAT 22
Db      766 AAAAAAAAAATTATAGATGAT 745
```

```
RESULT 47
US-10-750-185-34772/C
; Sequence 34772, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34772
; LENGTH: 1145
; TYPE: DNA
; ORGANISM: Bovine 19866880606879
US-10-750-185-34772
```

```
Query Match      68.8%; Score 17.2; DB 6; Length 1145;
Best Local Similarity 86.4%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTATAGCTTGAT 22
Db      625 AAAAAAAAACTTAACGTGAT 604
```

```
RESULT 48
US-10-750-185-26428
; Sequence 26428, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
```

```

; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26428
; LENGTH: 1270
; TYPE: DNA
; ORGANISM: Bovine 19866881425965
US-10-750-185-26428
```

```
Query Match      68.8%; Score 17.2; DB 6; Length 1270;
Best Local Similarity 86.4%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTATAGCTTGAT 22
Db      71 AAAAAAAAAATCATGTTGAT 92
```

```
RESULT 49
US-10-750-185-50582/C
; Sequence 50582, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50582
; LENGTH: 1425
; TYPE: DNA
; ORGANISM: Bovine 19866881244487
US-10-750-185-50582
```

```
Query Match      68.8%; Score 17.2; DB 6; Length 1425;
Best Local Similarity 86.4%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 AAAAAAAAACTATAGCTTGAT 22
Db      491 AAAAAAAAAAATGTTGAT 470
```

```
RESULT 50
US-10-750-185-24672
; Sequence 24672, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
```

```

; FILE REFERENCE: NM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24672
; LENGTH: 1433
; TYPE: DNA
; ORGANISM: Bovine 19866881220917
US-10-750-185-24672

```

```

Query Match      68.8%; Score 17.2; DB 6; Length 1433;
Best Local Similarity 86.4%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAACTATAGCTGAT 22
        ||||| ||||| ||||| |||||
Db      1325 AAAAAAGAAATCTATAGCTTCAT 1346

```

Search completed: December 14, 2005, 11:40:24
Job time : 186.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw.model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-5

Perfect score: 25
Sequence: 1 aaaaaaaaaagcactgactgtgacac 25

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_in: *
3: gb_env: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pr: *
9: gb_ro: *
10: gb_sts: *
11: gb_sy: *
12: gb_vl: *
13: gb_vl: *
14: gb_ncg: *
15: gb_pl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	93.6	41824	14	AC068427 Homo sapi
2	23.4	93.6	54810	14	AC027543 Homo sapi
3	23.4	93.6	153267	8	AC011212 Homo sapi
4	23.4	93.6	167272	8	AC011466 Homo sapi
5	22.4	89.6	2013	6	AX506824 Sequence
6	22.4	89.6	107527	15	FI9P19 AC000104 Sequence
7	22.4	89.6	161863	8	AL606970 Human DNA
8	22.4	89.6	213024	14	AC021185 Homo sapi
9	22.4	89.6	213024	14	AC163593 Bos tauri
10	21.8	87.2	132715	5	BX465189 Zebrafish
11	21.8	87.2	151034	14	AC156760 Dasyatis n
12	21.8	87.2	198036	5	BX004787 Zebrafish
13	21.8	87.2	222940	14	AC097839 Rattus no
14	21.8	87.2	240477	14	AC117921 Rattus no
15	21.8	84.0	116349	5	CR405685 Zebrafish
16	21.8	84.0	144165	5	AC002127 Homo sapi
17	21.8	84.0	168310	9	AC159480 Mus muscu
18	21.8	84.0	169542	8	AC146002 Pan trogl

19	21	84.0	173519	14	AC145860 AC145860 Pan trogl
20	21	84.0	175609	8	AC005243 AC005243 Homo sapi
21	21	84.0	182141	5	AL954645 Zebrafish
22	21	84.0	191529	8	AC161474 AC161474 Homo sapi
23	20.8	83.2	819	10	BV018408 BV018408 S212P6120
24	20.8	83.2	50277	14	AC138524_5 Continuation (6 of
25	20.8	83.2	110000	14	AC138524_4 Continuation (5 of
26	20.8	83.2	110000	14	BX308758_1 Continuation (12 of
27	20.8	83.2	110000	15	AP008210_312 Continuation (13 of
28	20.8	83.2	122652	5	CR391975 Zebrafish
29	20.8	83.2	133279	15	OSG1N00152 AL662950 Oryza sat
30	20.8	83.2	144093	9	AC109193 AC109193 Mus muscu
31	20.8	83.2	144375	14	CR354586 CR354586 Danio rer
32	20.8	83.2	146673	8	AC093755 AC093755 Homo sapi
33	20.8	83.2	153509	9	AC155334 AC155334 Mus muscu
34	20.8	83.2	166695	5	BX005050 BX005050 Zebrafish
35	20.8	83.2	180541	14	CR788236 CR788236 Danio rer
36	20.8	83.2	187045	14	AC073253 AC073253 Homo sapi
37	20.8	83.2	187819	14	AC137261 AC137261 Rattus no
38	20.8	83.2	191146	5	BX530070 BX530070 Zebrafish
39	20.8	83.2	194439	14	AC109022 AC109022 Rattus no
40	20.8	83.2	208833	14	AC161795 AC161795 Mus muscu
41	20.8	83.2	213656	14	AC161925 AC161925 Mus muscu
42	20.8	83.2	223824	14	AC121411 AC121411 Rattus no
43	20.8	83.2	225581	14	BX537105 BX537105 Danio rer
44	20.8	83.2	227705	14	AC152773 AC152773 Bos tauri
45	20.8	83.2	230172	14	AC120722 AC120722 Rattus no
46	20.8	83.2	236814	14	AC131436 AC131436 Rattus no
47	20.8	83.2	251434	14	AC162048 AC162048 Bos tauri
48	20.8	83.2	264197	14	AC123311 AC123311 Rattus no
49	20.8	83.2	270106	14	AC156526 AC156526 Bos tauri
50	20.8	83.2	280215	14	AC129648 AC129648 Rattus no
51	20.8	83.2	287236	14	AC156168 AC156168 Bos tauri
52	20.4	81.6	437	10	BV429670 BV429670 S237P6513
53	20.4	81.6	37931	8	AC005101 AC005101 Homo sapi
54	20.4	81.6	83705	8	AC025766 AC025766 Homo sapi
55	20.4	81.6	103925	9	AL607088 AL607088 Mouse DNA
56	20.4	81.6	141655	14	AC14894 AC14894 Felis cat
57	20.4	81.6	183972	8	AC093876 AC093876 Homo sapi
58	20.4	81.6	189303	14	AC148541 AC148541 Simiiri b
59	20.4	81.6	195558	8	AC007448 AC007448 Homo sapi
60	20.4	81.6	218909	14	AC018806 AC018806 Homo sapi
61	20.4	81.6	224215	14	AC114068 AC114068 Rattus no
62	20.4	81.6	232535	14	AC110702 AC110702 Rattus no
63	20.4	81.6	246162	14	AC133057 AC133057 Rattus no
64	20.4	81.6	262188	14	AC126973 AC126973 Rattus no
65	20.4	81.6	264977	14	AC097564 AC097564 Rattus no
66	20.4	81.6	283967	14	AC105589 AC105589 Rattus no
67	20.4	81.6	330608	14	AC109561 AC109561 Rattus no
68	20.2	80.8	426	10	BV387350 BV387350 S244P6341
69	20.2	80.8	654	10	BV275844 BV275844 S232P6233
70	20.2	80.8	680	10	BV499587 BV499587 S222P6644
71	20.2	80.8	707	10	BV472447 BV472447 G591P6170
72	20.2	80.8	749	10	BV609329 BV609329 S215P6549
73	20.2	80.8	898	10	BV465588 BV465588 G591P6612
74	20.2	80.8	945	15	AB11060 Epcatretu
75	20.2	80.8	95185	15	ATF617 CR936321 Zebrafish
76	20.2	80.8	95185	15	ATF617 AC007401 AC007401 Homo sapi
77	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
78	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
79	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
80	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
81	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
82	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
83	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
84	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
85	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
86	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
87	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
88	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
89	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
90	20.2	80.8	95185	15	ATF617 AL949657 Arabidops
91	20.2	80.8	95185	15	ATF617 AL949657 Arabidops

C 92	20.2	80.8	141048	14	AC166630	AC166630 Daaypus n
C 93	20.2	80.8	141584	14	AP005129	AP005129 Oryza sat
C 94	20.2	80.8	142411	14	AC148899	AC148899 Oryza sat
C 95	20.2	80.8	145016	8	AC123980	AC123980 Pan trogl
C 96	20.2	80.8	145476	14	AF050874	AF050874 Homo sapi
C 97	20.2	80.8	146066	9	AL935056	AL935056 Mouse DNA
C 98	20.2	80.8	148266	14	AC087532	AC087532 Homo sapi
C 99	20.2	80.8	149521	14	AC090903	AC090903 Homo sapi
C 100	20.2	80.8	150655	8	AC025887	AC025887 Homo sapi
C 101	20.2	80.8	151764	14	CR848709	CR848709 Danio rer
C 102	20.2	80.8	152732	15	OSG00168	OSG00168 Oryza sat
C 103	20.2	80.8	154893	14	AC072038	AC072038 Homo sapi
C 104	20.2	80.8	156035	5	BX005634	BX005634 Zebrafish
C 105	20.2	80.8	156578	8	AC102802	AC102802 Homo sapi
C 106	20.2	80.8	156939	5	CR354424	CR354424 Zebrafish
C 107	20.2	80.8	159413	14	AC155128	AC155128 Bos tauru
C 108	20.2	80.8	160955	14	AC132508	AC132508 Rattus no
C 109	20.2	80.8	161198	8	AF274854	AF274854 Homo sapi
C 110	20.2	80.8	163197	14	AC159368	AC159368 Bos tauru
C 111	20.2	80.8	163733	8	AC127526	AC127526 Homo sapi
C 112	20.2	80.8	166125	14	AC019327	AC019327 Homo sapi
C 113	20.2	80.8	166960	8	AC117513	AC117513 Homo sapi
C 114	20.2	80.8	168056	14	AC125167	AC125167 Mus muscu
C 115	20.2	80.8	173997	14	AC012340	AC012340 Homo sapi
C 116	20.2	80.8	174436	14	AC073076	AC073076 Homo sapi
C 117	20.2	80.8	174819	8	AC145688	AC145688 Pan trogl
C 118	20.2	80.8	174826	8	AC116562	AC116562 Homo sapi
C 119	20.2	80.8	176642	5	BX005226	BX005226 Zebrafish
C 120	20.2	80.8	177080	14	AC121090	AC121090 Mus muscu
C 121	20.2	80.8	177303	14	AC135998	AC135998 Homo sapi
C 122	20.2	80.8	177332	8	AC027290	AC027290 Homo sapi
C 123	20.2	80.8	179998	14	AC141350	AC141350 Rattus no
C 124	20.2	80.8	180728	14	AC164220	AC164220 Bos tauru
C 125	20.2	80.8	181432	14	AC134962	AC134962 Sus scrof
C 126	20.2	80.8	184042	8	AC011235	AC011235 Homo sapi
C 127	20.2	80.8	185441	9	AL824716	AL824716 Mouse DNA
C 128	20.2	80.8	186432	14	AC162617	AC162617 Bos tauru
C 129	20.2	80.8	186836	5	AL954677	AL954677 Zebrafish
C 130	20.2	80.8	188873	5	BX957260	BX957260 Zebrafish
C 131	20.2	80.8	190535	5	AC140305	AC140305 Mus muscu
C 132	20.2	80.8	190741	9	AC100381	AC100381 Mus muscu
C 133	20.2	80.8	191616	9	AC154619	AC154619 Mus muscu
C 134	20.2	80.8	193486	8	AC012123	AC012123 Homo sapi
C 135	20.2	80.8	197070	15	ATCHR1662	ATCHR1662 Arabidops
C 136	20.2	80.8	197545	14	AC135609	AC135609 Pan trogl
C 137	20.2	80.8	198191	9	AC108424	AC108424 Mus muscu
C 138	20.2	80.8	205588	8	AC068400	AC068400 Homo sapi
C 139	20.2	80.8	206478	9	AL928693	AL928693 Mouse DNA
C 140	20.2	80.8	208014	5	BX005008	BX005008 Zebrafish
C 141	20.2	80.8	208801	9	AC150560	AC150560 Mus muscu
C 142	20.2	80.8	211153	14	AC103436	AC103436 Rattus no
C 143	20.2	80.8	213782	14	AC011318	AC011318 Homo sapi
C 144	20.2	80.8	214145	5	BX511137	BX511137 Zebrafish
C 145	20.2	80.8	215166	9	AC117586	AC117586 Mus muscu
C 146	20.2	80.8	217604	14	AC152637	AC152637 Bos tauru
C 147	20.2	80.8	218258	9	AC096310	AC096310 Rattus no
C 148	20.2	80.8	220174	9	AC108852	AC108852 Mus muscu
C 149	20.2	80.8	221729	14	AC094575	AC094575 Rattus no
C 150	20.2	80.8	222594	14	AC124931	AC124931 Rattus no

ALIGNMENTS

RESULT 1
AC068427/c
LOCUS AC068427
DEFINITION Homo sapiens chromosome 12 clone RP11-252G6 map 12, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC068427
VERSION AC068427.1 GI:7677826
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 41824)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 12, clone RP11-252G6
JOURNAL
TITLE
AUTHORS
REFERENCE
REFERENCES
2 (bases 1 to 41824)
Unpublished

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F., Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquet, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galdan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N., Grant, G., Hages, B., Hestford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, U., Lacroque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPherson, R., Melgrim, J., Menues, L., Mihova, T., Miranda, C., Mianga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Roehman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigg, J., Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Center project name: L7012
Center clone name: 252_G_6
NOTE: This record contains 51 individual
contigs. Reads of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 725: contig of 725 bp in length
* 726 825: gap of 100 bp
* 826 1550: contig of 725 bp in length
* 1551 1650: gap of 100 bp
* 1651 2371: contig of 721 bp in length
* 2372 2471: gap of 100 bp
* 2472 3195: contig of 724 bp in length
* 3196 3295: gap of 100 bp
* 3296 4013: contig of 718 bp in length
* 4014 4113: gap of 100 bp
* 4114 4838: contig of 725 bp in length
* 4839 4938: gap of 100 bp
* 4939 5660: contig of 722 bp in length
* 5661 5760: gap of 100 bp
* 5761 6488: contig of 728 bp in length
* 6489 7308: gap of 100 bp in length
* 7308: contig of 720 bp in length

```

* 7309 7408: gap of 100 bp
* 7409 8080: contig of 672 bp in length
* 8081 8180: gap of 100 bp
* 8181 8911: contig of 731 bp in length
* 8912 9011: gap of 100 bp
* 9012 9741: contig of 730 bp in length
* 9742 9841: gap of 100 bp
* 9842 10565: contig of 724 bp in length
* 10566 10665: gap of 100 bp
* 10666 11404: contig of 739 bp in length
* 11405 11504: gap of 100 bp
* 11505 12208: contig of 704 bp in length
* 12209 12308: gap of 100 bp
* 12309 13034: contig of 726 bp in length
* 13035 13134: gap of 100 bp
* 13135 13853: contig of 719 bp in length
* 13854 13953: gap of 100 bp
* 13954 14699: contig of 746 bp in length
* 14700 14799: gap of 100 bp
* 14800 15523: contig of 724 bp in length
* 15524 15623: gap of 100 bp
* 15624 16392: contig of 769 bp in length
* 16393 16492: gap of 100 bp
* 16492 17224: contig of 732 bp in length
* 17225 17324: gap of 100 bp
* 17325 18048: contig of 724 bp in length
* 18049 18148: gap of 100 bp
* 18149 18877: contig of 729 bp in length
* 18878 18977: gap of 100 bp
* 18978 19710: contig of 733 bp in length
* 19711 19810: gap of 100 bp
* 19811 20549: contig of 735 bp in length
* 20550 20649: gap of 100 bp
* 20650 21361: contig of 712 bp in length
* 21362 21461: gap of 100 bp
* 21462 22168: contig of 707 bp in length
* 22169 22268: gap of 100 bp
* 22269 22971: contig of 703 bp in length
* 22972 23071: gap of 100 bp
* 23072 23803: contig of 732 bp in length
* 23804 23903: gap of 100 bp
* 23904 24579: contig of 676 bp in length
* 24580 24679: gap of 100 bp
* 24680 25405: contig of 727 bp in length
* 25407 25506: gap of 100 bp
* 25507 26229: contig of 723 bp in length
* 26230 26329: gap of 100 bp
* 26330 27067: contig of 738 bp in length
* 27068 27167: gap of 100 bp
* 27168 27888: contig of 721 bp in length
* 27889 27988: gap of 100 bp
* 27989 28687: contig of 699 bp in length
* 28688 28787: gap of 100 bp
* 28788 29500: contig of 713 bp in length
* 29501 29600: gap of 100 bp
* 29601 30277: contig of 677 bp in length
* 30278 30377: gap of 100 bp
* 30378 31110: contig of 733 bp in length
* 31111 31210: gap of 100 bp
* 31211 31938: contig of 728 bp in length
* 31939 32038: gap of 100 bp
* 32039 32752: contig of 714 bp in length
* 32753 32852: gap of 100 bp
* 32853 33572: contig of 720 bp in length
* 33573 33672: gap of 100 bp
* 33673 34391: contig of 719 bp in length
* 34392 34491: gap of 100 bp
* 34492 35224: contig of 733 bp in length
* 35225 35324: gap of 100 bp
* 35325 36054: contig of 730 bp in length
* 36055 36154: gap of 100 bp
* 36155 36883: contig of 729 bp in length
* 36884 36983: gap of 100 bp

```

```

* 36984 37701: contig of 718 bp in length
* 37702 37801: gap of 100 bp
* 37802 38512: contig of 711 bp in length
* 38513 38612: gap of 100 bp
* 38613 39340: contig of 728 bp in length
* 39341 39440: gap of 100 bp
* 39441 40178: contig of 738 bp in length
* 40179 40278: gap of 100 bp
* 40279 41007: contig of 729 bp in length
* 41008 41107: gap of 100 bp
* 41108 41824: contig of 717 bp in length.

FEATURES
source
1. .41824
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12"
/clone="RP11-252G6"
/clone_1lb="RP11 Human Male BAC"
726. .825
/estimated_length=100
1551. .1650
/estimated_length=100
2372. .2471
/estimated_length=100
3196. .3295
/estimated_length=100
4014. .4113
/estimated_length=100
4839. .4938
/estimated_length=100
5661. .5760
/estimated_length=100
6489. .6588
/estimated_length=100
7309. .7408
/estimated_length=100
8081. .8180
/estimated_length=100
8912. .9011
/estimated_length=100
9742. .9841
/estimated_length=100
10566. .10665
/estimated_length=100
11405. .11504
gap

Query Match 93.6%; Score 23.4; DB 14; Length 41824;
Best Local Similarity 96.0%; Pred. No. 20;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGACAGCAC 25
Db 39148 AAAAAAAAAAGCATGCTGACAC 39124

RESULT 2
AC027543 54810 bp DNA linear HTG 28-AUG-2002
DEFINITION Homo sapiens chromosome 19 clone RP11-601G24 map 19, LOW-PASS
SEQUENCE SAMPLING.
AC027543
AC027543.1 GI:7342288
VERSION HTG: HTGS PHASE0.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE 1 (bases 1 to 54810)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 19, clone RP11-601G24

```

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 54810)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Baetien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lechocky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPheters,R.,
Meldrum,J., Menues,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

JOURNAL
REFERENCE
AUTHORS
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 54810)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Baetien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lechocky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPheters,R.,
Meldrum,J., Menues,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

JOURNAL
REFERENCE
AUTHORS
Submitted (28-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L8832
Center clone name: 601_G_24

* NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

* be preserved.
1
708: contig of 708 bp in length
709
808: gap of 100 bp
809
1522: contig of 714 bp in length
1523
1622: gap of 100 bp
1623
2290: contig of 668 bp in length
2291
2390: gap of 100 bp
2391
3078: contig of 688 bp in length
3079
3178: gap of 100 bp
3179
3873: contig of 695 bp in length
3874
3973: gap of 100 bp
3974
4659: contig of 696 bp in length
4660
4769: gap of 100 bp
4770
5471: contig of 702 bp in length
5472
5571: gap of 100 bp
5572
6271: contig of 700 bp in length
6272
6371: gap of 100 bp
6372
7023: contig of 652 bp in length
7024
7123: gap of 100 bp
7124
7838: contig of 715 bp in length
7839
7938: gap of 100 bp
8624: contig of 686 bp in length
8625
8724: gap of 100 bp
8725
9392: contig of 668 bp in length
9393
9492: gap of 100 bp
9493
10198: contig of 706 bp in length
10199
10298: gap of 100 bp
10299
10192: contig of 714 bp in length
11013
11112: gap of 100 bp
11113
11807: contig of 695 bp in length
11808
11907: gap of 100 bp
11908
12611: contig of 704 bp in length
12612
12711: gap of 100 bp
12712
13419: contig of 708 bp in length
13420
13519: gap of 100 bp
13520
14187: contig of 668 bp in length
14188
14287: gap of 100 bp
14288
14998: contig of 711 bp in length
14999
15098: gap of 100 bp
15099
15808: contig of 710 bp in length
15809
15908: gap of 100 bp
15909
16587: contig of 679 bp in length
16588
16687: gap of 100 bp
16688
17374: contig of 687 bp in length
17375
17474: gap of 100 bp
17475
18167: contig of 693 bp in length
18168
18267: gap of 100 bp
18268
18870: contig of 703 bp in length
18871
19070: gap of 100 bp
19071
19782: contig of 712 bp in length
19783
19882: gap of 100 bp
19883
20528: contig of 646 bp in length
20529
20628: gap of 100 bp
20629
21331: contig of 703 bp in length
21332
21431: gap of 100 bp
21432
22119: contig of 688 bp in length
22120
22219: gap of 100 bp
22220
22913: contig of 694 bp in length
22914
23013: gap of 100 bp
23014
23574: contig of 661 bp in length
23575
23774: gap of 100 bp
23775
24470: contig of 696 bp in length
24471
24570: gap of 100 bp
24571
25267: contig of 697 bp in length
25268
25367: gap of 100 bp
25368
26073: contig of 706 bp in length
26074
26173: gap of 100 bp
26174
26876: contig of 703 bp in length
26877
26976: gap of 100 bp
26977
27662: contig of 686 bp in length
27663
27762: gap of 100 bp
27763
28438: contig of 676 bp in length
28439
28538: gap of 100 bp


```

* 28539 29267: contig of 729 bp in length
* 29368 29367: gap of 100 bp
* 29368 30076: contig of 709 bp in length
* 30077 30176: gap of 100 bp
* 30177 30832: contig of 656 bp in length
* 30833 30932: gap of 100 bp
* 30933 31599: contig of 667 bp in length
* 31600 31699: gap of 100 bp
* 31700 32390: contig of 691 bp in length
* 32391 32490: gap of 100 bp
* 32491 33188: contig of 698 bp in length
* 33189 33289: gap of 100 bp
* 33289 34002: contig of 714 bp in length
* 34003 34102: gap of 100 bp
* 34103 34801: contig of 699 bp in length
* 34802 34901: gap of 100 bp
* 34902 35595: contig of 694 bp in length
* 35596 36352: contig of 657 bp in length
* 36353 36452: gap of 100 bp
* 36453 37150: contig of 698 bp in length
* 37151 37250: gap of 100 bp
* 37251 37956: contig of 706 bp in length
* 37957 38056: gap of 100 bp
* 38057 38754: contig of 698 bp in length
* 38755 38854: gap of 100 bp
* 38855 39558: contig of 704 bp in length
* 39559 39658: gap of 100 bp
* 40379 40379: contig of 721 bp in length
* 40380 40479: gap of 100 bp
* 40480 41195: contig of 716 bp in length
* 41196 41295: gap of 100 bp
* 41296 41996: contig of 701 bp in length
* 41997 42096: gap of 100 bp
* 42097 42801: contig of 705 bp in length
* 42802 42901: gap of 100 bp
* 42902 43614: contig of 713 bp in length
* 43615 43714: gap of 100 bp
* 43715 44422: contig of 708 bp in length
* 44423 44522: gap of 100 bp

```

Query Match 93.6%; Score 23.4; DB 14; Length 54810;
 Best Local Similarity 96.0%; Pred. No. 19;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
 Db 10627 AAAAAAAAAAGCATGCTGTGACAC 10651

RESULT 3
 AC011212/c
 LOCUS
 DEFINITION Homo sapiens clone RP11-3N16, WORKING DRAFT SEQUENCE, 22 unordered
 pieces
 AC011212
 VERSION AC011212.3 GI:7230012
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 .Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 1 (bases 1 to 153267)
 Birren,B., Linton,L., Nussbaum,C. and Lander,E.
 Homo sapiens, clone RP11-3N16
 unpublished
 2 (bases 1 to 153267)
 Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Bouhgalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,

TITLE JOURNAL COMMENT

Galagan,J., Gardyna,S., Grant,G., Hages,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lechoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testaye,S., Tittrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (03-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 12, 2000 this sequence version replaced gi:6139137.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L2783
 Center clone name: 3 N 16
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 111839 bases at least Q40
 Consensus quality: 133540 bases at least Q30
 Consensus quality: 144810 bases at least Q20
 Insert size: 163000; agarose-fp
 Insert size: 151167; sum-of-contigs
 Quality coverage: 3.7 in Q20 bases; agarose-fp
 Quality coverage: 4.0 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 consists of 22 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1
* 1253 1252: contig of 1252 bp in length
* 1352 1352: gap of 100 bp
* 1353 2550: contig of 1198 bp in length
* 2551 2650: gap of 100 bp
* 2651 4705: contig of 2055 bp in length
* 4706 4805: gap of 100 bp
* 4806 7966: contig of 3161 bp in length
* 7967 8066: gap of 100 bp
* 8067 10438: contig of 2372 bp in length
* 10439 10538: gap of 100 bp
* 10539 13124: contig of 2586 bp in length
* 13125 13224: gap of 100 bp
* 13225 15586: contig of 2362 bp in length
* 15587 15686: gap of 100 bp
* 15687 19054: contig of 3368 bp in length
* 19055 19154: gap of 100 bp
* 19155 23442: contig of 4288 bp in length
* 23443 23542: gap of 100 bp
* 23543 28034: contig of 4492 bp in length
* 28035 28134: gap of 100 bp
* 28135 31029: contig of 2895 bp in length
* 31030 31129: gap of 100 bp
* 31130 36025: contig of 4886 bp in length
* 36026 36125: gap of 100 bp
* 36126 42963: contig of 6888 bp in length
* 42964 43063: gap of 100 bp
* 43064 51794: contig of 8721 bp in length
* 51795 51894: gap of 100 bp
* 51895 60991: contig of 9097 bp in length

```

```

*      60992      61091: gap of 100 bp
*      61092      70105: contig of 9014 bp in length
*      70106      70205: gap of 100 bp
*      70206      77517: contig of 7312 bp in length
*      77518      87761: gap of 100 bp
*      87762      87766: contig of 10109 bp in length
*      87767      87826: gap of 100 bp
*      87827      100723: contig of 12897 bp in length
*      100724      100823: gap of 100 bp
*      100824      115535: contig of 14712 bp in length
*      115536      115635: gap of 100 bp
*      115636      134718: contig of 19083 bp in length
*      134719      134818: gap of 100 bp
*      134819      153267: contig of 18449 bp in length.
*      153267      Location/Qualifiers
*      153267      /organism="Homo sapiens"
*      153267      /mol_type="genomic DNA"
*      153267      /db_xref="taxon:9606"
*      153267      /clone_id="RP11-3N16"
*      153267      /clone_lib="RP11-3N16 Human Male BAC"
*      153267      1..1252
*      153267      /note="assembly_fragment"
*      153267      1253..1352
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      1353..2550
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      2551..2650
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      2651..4705
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      4706..4805
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      4806..7966
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      7967..8066
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      8067..10438
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      10439..10538
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      10539..13124
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      13125..13224
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      13225..15586
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      15587..15686
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      15687..19054
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      19055..19154
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      19155..23442
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      23443..23542
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      23543..28034
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      28035..28134
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      28135..31029
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      31030..31129
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      31130..36025
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      36026..36125
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      36126..42963
*      153267      /estimated_length=100
*      153267      gap
*      153267      misc_feature
*      153267      /note="assembly_fragment"
*      153267      42964..43063
*      153267      /estimated_length=100
*      153267      gap

```

```

misc_feature      43064..51794
                  /note="assembly_fragment"
gap              51795..51894
                  /estimated_length=100
misc_feature      51895..60991
                  /note="assembly_fragment"
gap              60992..61091
                  /estimated_length=100
misc_feature      61092..70105
                  /note="assembly_fragment"
gap              70106..70205
                  /estimated_length=100
misc_feature      70206..77517
                  /note="assembly_fragment"
gap              77518..87761
                  /note="assembly_fragment"
misc_feature      87762..87766
                  /estimated_length=100
gap              87767..87826
                  /estimated_length=100
misc_feature      87827..100723
                  /estimated_length=100
gap              100724..100823
                  /estimated_length=100
misc_feature      100824..115535
                  /note="assembly_fragment"
gap              115536..115635
                  /estimated_length=100
misc_feature      115636..134718
                  /estimated_length=100

```

```

Query Match      93.6%; Score 23.4; DB 14; Length 153267;
Best Local Similarity 96.0%; Pred. No. 16;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGACTGTGACAC 25
DB      92634 AAAAAAAAAAGCATGACTGTGACAC 92610

RESULT 4
AC011466      167272 bp      DNA      linear      PRI 31-JUL-2001
LOCUS      Homo sapiens chromosome 19 clone CTC-45323, complete sequence.
DEFINITION      AC011466
ACCESSION      AC011466.6      GI:15042800
VERSION      HTG.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 167272)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 167272)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jul 31, 2001 this sequence version replaced gi:1369565.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center

```


CDS
 complement(join(2612..2739,2831..2901,2964..3087,
 3227..3311,3424..3531,3581..3700,4248..4307,4403..4498,
 4757..4873,5037..5158,5205..5421,5538..5624,5791..5826,
 5895..6000,6046..6212,6267..6352,6443..6455))
 /gene="F19p19.2"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAB70423.1"
 /db_xref="GI:2341025"
 /translation="NMCAISGEVPEEPVVSKKSGLLYERKLIOTHTISLFLSLHTSIC
 RYELNDFVILIRLLVPCRIKMGARLLVSHLLMTLFPKSGRLFKPDSISGTL
 OLVKPELHDASIPGLGTFOTDHIETKXNNILWMDSLMNSPALBEOQLHARAL
 HALVYIDGVTPEPLDHAACRYIARLKRRDSRRLABERQLPAEVAVTSNAL
 SNKGMRLRLINLWLCGIDGECGPNAKKRIGISAIVTETLDCNAALSOQKK
 ROIPTLASVDALKEFTQLSHPILKTKNPGIFSMIDILSKVIAVGGIDTAVIDR
 PSGQLISTLTHGSKVTSIKFVGDIDLVLTASDENDVNTAAAPDPGLIL
 GYTAOSIVIKIMDVKSQFVLTSLSLSFSDROMQSLVDWMEKSLYSLKTTALD
 GRLMDRLKXNRTFDPDANGGEVPTQCEMLIDIGAHDPNEILKDSGANSINIL
 SPBMQGVFOAASVKAENNPITKLPLDSTGKATSVKFGLDKTYIAVGSMDNRIRFG
 LPDDNTBDSAQDS"
 /protein_id="F19p19.3"
 /db_xref="GI:2341026"
 /translation="METLPVPLFKYIVSCFTRKSHQLAFYASLQMSINKOLDEIT
 SPTRADIFPDCSOLFQRTFNQTSBVTASNGCYENNNNTNFPDKSNGSQDIED
 NNDNALSLIFDSQDPNDITASIDFSSSIOPASDQLQEQDFTGILHOPNTLY
 SSSGSGILPPLSVFEDCLSVPSINSIPSCSFLGNTGLPTMTYPMNMN
 TIGSGFSGNHLIGSDPKFSDHQLMEIOANGLCFPPKIPFNPQHHLOGLGV
 EKNONHVAQVPLQGLTEITGLDDPSFNKVGKLSARQEKIHRVKKNNBNSFKI
 KYPLVHNHRIRKPYACRKTILADSRPRVGRPAKNDPFGEPNRAQCSHHEDDDIV
 LILFDYN"
 /gene="F19p19.4"
 /join(13879..14922,15023..15175,15228..15269)
 /gene="F19p19.4"
 /note="EST gb|21788 comes from this gene."
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAB70425.1"
 /db_xref="GI:2341027"
 /translation="MEGOTHDQNNVBEHDNPLMVKNPSSNLFSKKAQCLSYLQNG
 KMLQVRSSPKVSCEASTCPSHITLVGENVYAEKMKKXVSPSKTPSGQLOFT
 MANGTPHVPFKLEKQDYVASLSNVODNSYTHLORGBSSASSSHLVGINST
 LFSKVLREFFVLFSNGENLKIPTKRRGSKKVAVHAVKERRARLSRTSPIDL
 GSWDEQFOAONDCLLKNKLPTNLFTLAVVQETIEDBIGGGLFKLRSFMPFORN
 DASETSTSSISMNVIPSGIHGDEDPSSILIEKSGQNGCDLCSLTLLK
 GQPRKQVFEFLIEGSKHETGLKIVNVSGLYLVQFEAKLTSLSGFALALAFHSEK
 LRPILHLYLEKQILERRSP"
 /complement(15947..17024)
 /gene="F19p19.5"
 /complement(join(15947..16071,16676..17024))
 /gene="F19p19.5"
 /note="Strong similarity to 60S ribosomal protein L17
 (gb|X01694). EST gb|AA042332 comes from this gene."
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAB70426.1"
 /db_xref="GI:2341028"
 /translation="MCLFSFSNLVILVVDLDFRLGRGSGNKFMSGLPYAATVN
 CADNTGAKNLYITISVKGIRGLNRLPSCACGMVMAVTVKRGDLAKKYLPAVIYQR
 KPMRRKQGVPMYFEDNAGVIVNPKGSKSATIGPICKCADLMPRIASAAAIIV"
 /gene="F19p19.6"
 /join(18613..19224,19307..19974,20046..20718,20796..20982,
 21060..22027)
 /gene="F19p19.6"

/note="EST gb|ATTS5672 comes from this gene."
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAB70427.1"
 /db_xref="GI:1903347"
 /translation="MVRHSRRSPFDPMVGFVCPPDILMPKGLDGLDRDRREAY
 EIFPAACSSPGFGRNALTFYSKRNAGDHQDGGGGGSSSSNSGSGSLARKEVL
 TPTFSVRKALGLKMLKSPSRMSTVGVVAVAPSPSFGNNGSIGSSGHPSPAG
 PFTVPPARRRRLTSALIEIRQMKVTSQDILRKTLMTLVQGTGRARAEIILPLEL
 LHVHVSBERGDVHEVQIMQRLQVLEBALIHPSPLEKTNVFAARLAEIIRQSETK
 AIDTKNSDIMPTELNLVASSLRNATPPTDICHMADGYPLNHLVALLOSIFDIRD
 ETLVLDIDELIELMKTMIMGITRAIHLNLFMTVLPQVYVTSQMEBIDLGASAM
 LAEVANDAKSREARLYVULSTLASMGRTBEKULLSHDYFORNVGILNLEPLA
 LSSSKILGEDVITISQNGLEKGDVLDVSGRVDTYIRASIKNAPSKITENKABIE
 ETEBEEBAATMLRLAKETEDLARESECFPIIKRMHLVAGVAVSLHOCYGSIL
 MOYLAGESTTETEVEVLOAGLEKLELVQVMAENSDECEGKGLVREMPVEVDIS
 ILRLRWMEKLOVVOELGSLAKEAEFTNPPSKSEPYVSGELMKLANDAIEEPE
 IPGIDTEDLVHLAQLBKLPQEVYTFVASCOSKOSYPTLPFLRCNDSKPKLWK
 KATPCASGEELNMGEPAGNHPPTSRGTORLYIRNTLHFLSGLHSLKSLSL
 NPRVLPATRKRCRERTKSSYPEFTQAGIESACQHVSAARLILFDLSVYFBSLY
 PGDVANGRIKPRALIKONLTMTAILADKAQALAMKEVMKASFVLLVTLAAGHSR
 VFCRTDHDILIEDFSLKVCCTCGGLIPESVVRBAETVGVIOLMQPTBOLMED
 FSLVCESSGMGLVGTGQLPMPPTTGRNRDPTILKVLGYRDDRVNQFLKKSFO
 LGRB"
 /gene="F19p19.7"
 /join(22508..22520,22653..22885)
 /gene="F19p19.7"
 /note="Strong similarity to 50S ribosomal protein L28
 (gb|X68078)."
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAB70428.1"
 /db_xref="GI:1903348"
 /translation="MLPBRICPFPGKANKRANKVPSNHTKKQLQFVNLQYKRVMB
 AKRRVILRLSKALTKTEKNGLEAVSKAAGIDLKRX"
 /complement(23351..24198)
 /gene="F19p19.8"
 /complement(join(23351..23802,24144..24198))
 /gene="F19p19.8"
 /codon_start=1
 /evidence=not_experimental
 /product="F19p19.8"

Query Match 89.6%; Score 22.4; DB 15; Length 107527;
 Best Local Similarity 95.8%; Pred. No. 44;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGACA 24
 Db 32791 AAAAAACAAGCATGACTGTGACA 32814

RESULT 7
 AL606970/c
 LOCUS
 DEFINITION
 Human DNA sequence from clone RP11-503C24 on chromosome 6 Contains
 the 3' end of a novel gene, up to five novel genes, (including
 FLJ31232 and p13671), a meningioma expressed antigen 6
 (coiled-coil proline-rich) (MGEA6) pseudogene and three Cpg
 islands, complete sequence.
 ACCESSION
 AL606970
 VERSION
 AL606970.12 GI:17065766
 KEYWORDS
 HTG; Cpg Island; FLJ31232; MGEA6; p13671.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominoidea; Homo.
 1 (bases 1 to 161863)
 REFERENCE
 AUTHORS
 Tracey, A.
 TITLE
 Direct Submision
 JOURNAL
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

COMMENT

Cambridgehire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Nov 25, 2001 this sequence version replaced gi:17017816.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Mp, MORMEP; Information on the MORMEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
Rp11-503C24 is from the library RPICT-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

SOURCE

Location/Qualifiers
1. 161863

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-503C24"
/clone_lib="RPICT-11.2"
1

misc_feature

/note="Clone_left_end: RP11-503C24"
join(complement(Al611929.2:9197. .9527), 2500. .26117)
/locus_tag="RP11-503C24.1-001"
join(complement(Al611929.2:9197. .9527), 2500. .2617)
/locus_tag="RP11-503C24.1-001"

gene

/note="match: ESTs: A1149978"
complement(join(23395. .23540, 24425. .24633))
/locus_tag="RP11-503C24.2-001"
complement(join(23395. .23540, 24425. .24633))
/locus_tag="RP11-503C24.2-001"

mRNA

/note="match: ESTs: A1825113 AW295767"
join(27848. .27970, 37138. .37225, 38570. .38831)
/locus_tag="RP11-503C24.3-001"
join(27848. .27970, 37138. .37225, 38570. .38831)
/locus_tag="RP11-503C24.3-001"

misc_feature

/note="match: ESTs: A1829659"
33700. .33768
/locus_tag="RP11-503C24.3-001"
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

misc_feature

34333. .34503
/locus_tag="RP11-503C24.3-001"
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

misc_feature

34695
/locus_tag="RP11-503C24.3-001"
/note="Random repeat. Forced join. Gap size estimated to be approximately 500bp by restriction digest data."

misc_feature

34826. .35086
/locus_tag="RP11-503C24.3-001"
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

gene

join(41054. .41191, 56984. .57492, 57781. .58006, 58784. .59486, 60028. .60697)
/locus_tag="RP11-503C24.4-001"
join(41054. .41191, 56984. .57492, 57781. .58006, 58784. .59486, 60028. .60697)

mRNA

/locus_tag="RP11-503C24.4-001"
/note="match: CDNAs: AK055794"
68376

misc_feature

/note="Tandem repeat. Forced join. Gap size estimated to be approximately 300bp by restriction digest data."

gene

complement(85021. .87466)
/locus_tag="RP11-503C24.5-001"
/note="match: proteins: O15320 Q95046 Q96R76 Q96SG9"

CDS

/pseudo
/codon_start=1
complement(join(98299. .98705, 100193. .100230))
/locus_tag="RP11-503C24.6-001"

gene

complement(join(98299. .98705, 100193. .100230))
/locus_tag="RP11-503C24.6-001"
/note="match: ESTs: BG428602"

mRNA

complement(105020. .117838)
/locus_tag="RP11-503C24.7-001"
complement(join(105020. .107214, 108284. .108562, 109268. .109400, 117504. .117838))
/gene="RP11-503C24.7"

gene

/locus_tag="RP11-503C24.7-001"
/note="match: ESTs: AW150833 AW763170 BB638300 BB651889 B1052777
match: CDNAs: AF318336"

polya_site

complement(105020)
/gene="RP11-503C24.7"
/locus_tag="RP11-503C24.7-001"
complement(105046. .105051)
/gene="RP11-503C24.7"

polya_signal

/locus_tag="RP11-503C24.7-001"
complement(join(105548. .107214, 108284. .108562, 109268. .109400, 117504. .117749))
/gene="RP11-503C24.7"

CDS

/locus_tag="RP11-503C24.7-001"
/standard_name="OTTHUMP00000017658"
/note="match: proteins: Q8WYM2"
/codon_start=1
/protein_id="CA114850.1"
/db_xref="GI:55962022"

misc_feature

/translacion="MWTGCGPPGSGACMDRRRLGARLRAPAGLQELGLRATQOERVR GALALPPAPAPACGPHGLHGEQOLEALALLOLSRLRODGLKTHLOLDL ISLLOLDVGAEGEALDSDSPSSGFEVMSDGGSCSLTSCAVCSGDKTSPISGLIP VAQAHKARSRSMEDWRRSYDVTTPVAPRQATBEGRPPRESVEDAQPACTTPRPVTS TGDLDLPAUDTGLQKASADAEILGLCCQVDIPLVHPDKYQDVLVSQGRVAVYVP SP1HAVALQSPLEFLTKETPQRGSPRPRESPPGAPLNTIQGPVLEAGPARAVAT DRLIHLMGRETPAKSGEGEGQPLRHAASPOQCGMSTGCGRLVFAGRDEGAP AGRGAGCGPQOOGYMPLEGPOGSGLEPEGKSPNSCYLRFTWVOASPSKAOQRP SADDYKRLNIDSPBSMLDKSPRSAGHPHAPSSGLKMGPPSVKXKTKRSMDYTL RPARQGLNIDPREGHNAPOSLBENDPNAHPTGREGLOKRLANLBAPEKSGSESTL YPMFVLVPLAVAPQESHRTSAQLPFPEASLTVARRRKRRQSTVEISARARLASC PESSNLCPPPPVARRAGCPPLARGPSLVRODAYRSDSPSKHACDPPRPVPIPERTS BGSSSDHTNRFCDDRESSSDREGAGQSDCDLGVAAAGHAEIAMTQAPVSSGPI LSPVPLCRITKASKALKKKIRRPQPALKMTWV"

misc_feature

112915
/note="Clone_right_end: RP11-3C9"

misc_feature

115621
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."

misc_feature

159864
/note="Clone_left_end: RP1-3J52"

misc_feature

161496
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 420bp by restriction digest data."

misc_feature

161496
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 420bp by restriction digest data."

misc_feature

161496
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 420bp by restriction digest data."

misc_feature

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 161863;
 Best Local Similarity 95.8%; Pred. No. 42;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24

Db 135962 AAAAAAAAAAGCATGACTGTGACA 135939

RESULT 8

LOCUS AC021185 201611 bp DNA linear HTG 20-OCT-2001
 DEFINITION Homo sapiens chromosome 6 clone RP11-73815, WORKING DRAFT SEQUENCE,
 5 unordered pieces.

AC021185 AC021185.4 GI:16259198

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Homo.

1 (bases 1 to 201611)

Waterston,R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 201611)

Waterston,R.H.

Submitted (14-JAN-2000)

Direct Submission

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Oct 18, 2001 this sequence version replaced gi:10944513.

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

Contact: submissions@wustl.edu

----- Project Information -----

Center project name: H_NH0738105

----- Summary Statistics -----

Sequencing vector: M13; 35%

Chemistry: Dye-terminator Big Dye; 65% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 201095 bases at least Q40

Consensus quality: 201923 bases at least Q30

Consensus quality: 202374 bases at least Q20

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 16969: contig of 16969 bp in length
 * 16970 17069: gap of unknown length
 * 17070 33654: contig of 16585 bp in length
 * 33655 33754: gap of unknown length
 * 33755 53582: contig of 19828 bp in length
 * 53583 53682: gap of unknown length
 * 53683 126590: contig of 72908 bp in length
 * 126591 201611: gap of unknown length
 * 126691 201611: contig of 74921 bp in length.

Location/Qualifiers

1. .201611

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

ORIGIN

Query Match 89.6%; Score 22.4; DB 14; Length 201611;
 Best Local Similarity 95.8%; Pred. No. 41;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24

Db 42259 AAAAAAAAAAGCATGACTGTGACA 42282

RESULT 9

AC163593/c

LOCUS

DEFINITION

unorderd pieces.

AC163593

VERSION

KEYWORDS

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;

Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 213024)

Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alshrock, S., Amlin, A., Anguiano, D.,

Anyalebech, V., Ayagi, A., Ayodeji, M., Baca, B., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Buttrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Crease, A., D'Souza, L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Davies, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregiorgis, R., Geer, K., Gill, R., Gizard, M., Guerra, M., Guevara, W.,

Gunnarathne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,

Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,

Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovar, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenz, H., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapus, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muniadasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, G., Nguyen, N., Norris, S., Nwokenkeme, O., Okunolu, G., Olarunpogon, A., Pal, S., Parks, K., Paaternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindecker, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puzo, M., Quito, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Swatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, V., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Neidhauser, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Unpublished
Direct Submission
2 (bases 1 to 213024)
Worley, K.C.
Direct Submission
Submitted (13-JUN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 213024)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2005 this sequence version replaced gi:67514866.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: FHTB
Center clone name: CH240-12515
----- Summary Statistics
Assembly program: Atlas 3.0:
Consensus quality: 203030 bases at least Q40
Consensus quality: 205410 bases at least Q30
Consensus quality: 207269 bases at least Q20
Estimated insert size: 206505; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES	source
1	24745: contig of 24745 bp in length
2	24746: gap of 545 bp
3	25291: contig of 32492 bp in length
4	57783: gap of 77 bp in length
5	57855: gap of 7161 bp in length
6	57860: contig of 186 bp
7	65207: gap of 38911 bp in length
8	104118: gap of 50 bp
9	104168: contig of 28467 bp in length
10	132635: gap of 50 bp
11	132685: contig of 10368 bp in length
12	143053: gap of 50 bp
13	143103: contig of 7536 bp in length
14	150639: gap of 50 bp
15	150689: contig of 5950 bp in length
16	156639: gap of 50 bp
17	156689: contig of 13081 bp in length
18	169770: gap of 50 bp
19	169819: contig of 4302 bp in length
20	174122: gap of 50 bp
21	174172: contig of 10642 bp in length
22	184814: gap of 50 bp
23	184864: contig of 6493 bp in length
24	191357: gap of 50 bp
25	191407: contig of 3754 bp in length
26	195161: gap of 349 bp
27	195510: contig of 3254 bp in length
28	198764: gap of unknown length
29	198863: gap of 1310 bp in length
30	200174: gap of unknown length
31	200274: contig of 1587 bp in length
32	201860: gap of unknown length
33	201960: contig of 1270 bp in length
34	201961: gap of unknown length
35	203231: contig of 1024 bp in length
36	203331: gap of unknown length
37	204355: contig of 1048 bp in length
38	204455: gap of unknown length
39	205503: contig of 1452 bp in length
40	205603: gap of unknown length
41	207055: contig of 1097 bp in length
42	207154: gap of unknown length
43	208251: contig of 1097 bp in length
44	208252: gap of unknown length
45	208352: contig of 4673 bp in length.
46	Location/Qualifiers
47	1. .213024
48	/organism="Bos taurus"
49	/mol_type="genomic DNA"
50	/db_xref="taxon:9913"
51	/clone="CH240-12515"
52	24746..25290
53	/estimated_length=545
54	57783..57859
55	/estimated_length=77
56	65021..65206
57	/estimated_length=186
58	104118..104167
59	/estimated_length=50
60	132635..132684
61	/estimated_length=50
62	143053..143102
63	/estimated_length=50
64	150639..150688
65	/estimated_length=50
66	156639..156688
67	/estimated_length=50

```

gap      169770..169819
          /estimated_length=50
gap      174122..174171
          /estimated_length=50
gap      184814..184863
          /estimated_length=50
gap      191357..191406
          /estimated_length=50
gap      195161..195199
          /estimated_length=349
gap      198764..198863
          /estimated_length=unknown
gap      200174..200273
          /estimated_length=unknown
gap      201861..201960
          /estimated_length=unknown
gap      203231..203330
          /estimated_length=unknown
gap      204355..204454
          /estimated_length=unknown
gap      205503..205602
          /estimated_length=unknown

```

```

Query Match      89.6%; Score 22.4; DB 14; Length 213024;
Best Local Similarity 95.8%; Pred. No. 40;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAGCATGACTGTGACA 24
DB      58879 AAAAAAAAAAGCATGACTGTGACA 58856

RESULT 10
BX465189      132715 bp      DNA      linear      VRT 14-JUL-2004
LOCUS      Zebrafish DNA sequence from clone RP71-47G15 in linkage group 23,
DEFINITION      complete sequence.
ACCESSION      BX465189
VERSION      BX465189.9 GI:50299987
KEYWORDS      HTG.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
            1 (bases 1 to 132715)
            BIlwood,M.
REFERENCE      Submitted (14-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
AUTHORS      Cambridgehire, CB10 1SA, UK. E-mail enquiries:
TITLE      zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
JOURNAL      On Jul 14, 2004 this sequence version replaced gi:50284456.
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Emi, EMBL, Swi, SWISSPROT, Tr, TREMBL, Wp, WORMPEP, Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted. Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmark.shtml RP71-47G15 is from a Zebrafish mixed sex BAC library VECTOR: pTARBAC2.

```

FEATURES
    source
        1..132715
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="RP71-47G15"
            /clone_11b="RPCT-71"

```

```

Query Match      87.2%; Score 21.8; DB 5; Length 132715;
Best Local Similarity 92.0%; Pred. No. 77;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAGCATGACTGTGACAC 25
DB      131315 AAAAAAAAAAGCATGACTGTAAAC 131339

RESULT 11
AC156760      151034 bp      DNA      linear      HTG 17-MAR-2005
LOCUS      Dasyus novemclinctus clone VMRC5-21118, WORKING DRAFT SEQUENCE, 6
DEFINITION      ordered pieces.
ACCESSION      AC156760
VERSION      AC156760.2 GI:61562514
KEYWORDS      HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Dasyus novemclinctus (nine-banded armadillo)
ORGANISM      Dasyus novemclinctus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Xenarthra; Dasypodidae; Dasypus.
            1 (bases 1 to 151034)
            Antonellis,A., Ayala,K., Benjamin,B., Blakeley,R.W., Boake,A.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H.,
            Engle,J., Gesole,M., Greene,A., Guan,X., Gupta,J., Haghighi,P.,
            Han,J., Hansen,N., Ho,S.-L., Hu,P., Hunter,G., Hurle,B., Idol,J.R.,
            Kwong,P., Latic,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R.,
            Madden,M., Maduro,Q.L., Maduro,V.B., Margulies,B.H., Masello,C.,
            Masferrer,B., McDowell,J., Mojidi,H.A., Mullikin,J.C.,
            Oestreicher,J.S., Park,M., Portnoy,M.B., Prasad,J.C., Putl,O.,
            Reddix-Jugue,N., Schandler,K., Schueler,M.G., Stilson,C.,
            Stantirpop,S., Stephen,E., Teye,A., Thomas,J.W., Thomas,P.J.,
            Tsipouri,V., Ung,U., Vogt,J.L., Weherby,K.D., Young,A. and
            Green,E.D.
REFERENCE      NISC Comparative Sequencing Initiative
AUTHORS      Unpublished
            2 (bases 1 to 151034)
            Green,E.D.
JOURNAL      Submitted (03-FEB-2005) NIH Intramural Sequencing Center, 5625
TITLE      Fishers Lane, Rockville, MD 20852, USA
JOURNAL      3 (bases 1 to 151034)
            Green,E.D.
REFERENCE      Submitted (17-MAR-2005) NIH Intramural Sequencing Center, 5625
AUTHORS      Fishers Lane, Rockville, MD 20852, USA
JOURNAL      On Mar 17, 2005 this sequence version replaced gi:58531485.
TITLE      ----- Genome Center
COMMENT      Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Contact: misc.zoo@hngri.nih.gov
 ----- Project Information
 Center project name: Kai
 Center clone name: 021118

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is generally based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 149393 bases at least Q40
 Consensus quality: 149999 bases at least Q30
 Consensus quality: 150398 bases at least Q20
 Insert size: 136000; agarose-fp
 Insert size: 150534; sum-of-contigs
 Quality coverage: 11.25x in Q20 bases; agarose-fp
 Quality coverage: 10.16x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 3279: contig of 3378 bp in length
 * 3279 3378: gap of unknown length
 * 3379 30684: contig of 27306 bp in length
 * 30685 30785: gap of unknown length
 * 30785 75510: contig of 44726 bp in length
 * 75511 75610: gap of unknown length
 * 75611 83080: contig of 7470 bp in length
 * 83081 83180: gap of unknown length
 * 83181 103803: contig of 20622 bp in length
 * 103803 103902: gap of unknown length
 * 103903 151034: contig of 47132 bp in length.
 * Location/Qualifiers
 1. 151034
 /organism="Darypus novemcinctus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9361"
 /clone="VMRC5-21118"
 /clone_1db="VMRC5"
 /note="BAC resource: http://bacpac.chori.org/"
 1. 100826
 /note="clone overlaps with GenBank Accession Number
 AC154942 clone VMRC5-359L24 (center project name ivo)"
 1. 3278
 /note="assembly_fragment
 clone end: T7
 vector_side: left"
 3279. 3378
 /estimated_length=unknown
 3379. 30684
 /note="assembly_fragment"
 30685. 30784
 /estimated_length=unknown
 30785. 75510
 /note="assembly_fragment"
 75511. 75610
 gap

misc_feature /estimated_length=unknown
 75611..83080
 /note="assembly_fragment"
 gap 83081..83180
 /estimated_length=unknown
 misc_feature 83181..103802
 /note="assembly_fragment"
 gap 103803..103902
 /estimated_length=unknown
 misc_feature 103903..151034
 /note="assembly_fragment
 clone end:SP6
 vector_side:right"
 131519..151034
 /note="clone overlaps with GenBank Accession Number
 AC155795 clone VMRC5-504M24 (center project name ivo)"

ORIGIN
 Query Match 87.2%; Score 21.8; DB 14; Length 151034;
 Best Local Similarity 92.0%; Pred. No. 75;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
 Db 126917 AAAAAAAAAAGAACTGACAC 126941

RESULT 12
 LOCUS BX004787 198036 bp DNA linear VRT 22-JUL-2003
 DEFINITION Zebrafish DNA sequence from clone DKEY-28B17 in linkage group 23,
 complete sequence.
 ACCESSION BX004787
 VERSION BX004787.8 GI:33146025
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 198036)
 Whitehead,S.
 Direct Submission
 Submitted (22-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 22, 2003 this sequence version replaced gi:32398517.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormpep/Clone-derived
 Zebrafish pUC subclones occasionally display inconsistency over the

length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
 Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmark.shtml DKEY-28B17 is from a Zebrafish BAC library
 VECTOR: pindigBAC-5.

FEATURES

SOURCE

Location/Qualifiers
 1. 198036
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEY-28B17"
 /clone_lib="DanioKey"

ORIGIN

Query Match 87.2%; Score 21.8; DB 5; Length 198036;
 Best Local Similarity 92.0%; Pred. No. 72;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACGTGACAC 25
 |||||
 Db 600 AAAAAAAAAAGCATGACGTGAAAC 624

RESULT 13

AC097839/c

LOCUS

AC097839 222940 bp DNA linear HTG 15-NOV-2002
 Rattus norvegicus clone CH230-178M15, *** SEQUENCING IN PROGRESS

DEFINITION

AC097839 AC097839.6 GI:25012184
 HTG: HTGS PHASE2: HTGS_DRAFT: HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Rattus norvegicus
 Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.
 1 (bases 1 to 222940)

REFERENCE

AUTHORS

Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alibonco, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F., Biewick, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Dlyva, K., Draper, H., Dugan-Roches, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C., Gabiel, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hui, Y., Hume, J., Idelberg, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowitz, C., Kraft, C., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzen, L., Loulsged, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martin, E., Mawhney, S., McLeod, M., McNeill, T., Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munida, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Nwackeme, O., Okunnu, G., Olarinmoye, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plapper, F., Polindere, A., Popovic, D., Prims, E., Pu, L., L., Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ritz, S., S., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C., D., Smales, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Soe, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K., Valsar, V., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wiczyska, R., Wooten, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 222940)
 Worley, K. C.
 Direct Submission
 Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 222940)
 Rat Genome Sequencing Consortium.
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:2366327.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GFJA
 Center clone name: CH230-178M15
 ----- Summary Statistics
 Assembly program: Phrap, version 0.990329
 Consensus quality: 202863 bases at least Q40
 Consensus quality: 206220 bases at least Q30
 Consensus quality: 207937 bases at least Q20
 Estimated insert size: 208748; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
 ----- NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_drafile.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 1 222940: contig of 222940 bp in length.

```

FEATURES
  source
    1. 222940
      /location/Qualifiers
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10116"
      /clone="CH230-178M15"
      1. 2455
      /note="wgs contig"
  misc_feature
    30328..31596
      /note="wgs_contig"
  misc_feature
    60543..61949
      /note="wgs_contig"
  ORIGIN

Query Match      87.2%; Score 21.8; DB 14; Length 222940;
Best Local Similarity 92.0%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 AAAAAAAAAAGCATGCTGACAC 25
Db      63487 AAAAAAAAAAGCATGCTGACGC 63463

RESULT 14
AC117921
LOCUS
DEFINITION
  Rattus norvegicus clone CH230-130U10, WORKING DRAFT SEQUENCE, 5
  unordered pieces.
AC117921
AC117921.6 GI:24941717
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULUTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurongnathi; Murioidea; Muridae; Murinae; Rattus.
  1 (bases 1 to 240477)
  Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,
  Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
  Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
  Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
  Biewald,K, Blair,J, Blankenburg,K, Blythe,P, Brown,M,
  Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
  Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
  Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
  Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
  Devila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
  Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
  Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
  Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falle,T, Fan,G,
  Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
  Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
  Gebregregis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W,
  Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
  Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
  Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
  Hollins,B, Howell,S, Huily,S, Hume,J, Idlebird,D, Jackson,A,
  Jackson,B, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
  Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
  Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
  Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
  Lorenshuwa,L, Louieged,H, Lozada,R,J, Lu,X, Ma,J,
  Maheswari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A,
  Mangum,B, Mapa,P, Martin,K, Martin,R, Martineau,E,
  Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
  Mlloasvilevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
  Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L,
  Menkervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
  Nnaekolehen,O, Okwomou,G, Olarinpunsagoon,A, Pal,S, Parks,K,
  Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C,
  Plopper,F, Poldexter,A, Popovic,E, Primus,E, Pu,L,-L,
  Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
  Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richard,S, Riggs,F,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
  Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  3 (bases 1 to 240477)
  Rat Genome Sequencing Consortium.
  Direct Submission
  Submitted (11-APR-2002) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  3 (bases 1 to 240477)
  Rat Genome Sequencing Consortium.
  Direct Submission
  Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  On Nov 13, 2002 this sequence version replaced gi:23194988.
  The sequence in this assembly is a combination of BAC based reads
  and whole genome shotgun sequencing reads assembled using Atlas
  (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
  in the feature table below represents a scaffold in the Atlas
  assembly (a 'contig-scaffold'). Within each contig-scaffold,
  individual sequence contigs are ordered and oriented, and separated
  by sized gaps filled with Ns to the estimated size. The sequence
  may extend beyond the ends of the clone and there may be sequence
  contigs within a contig-scaffold that consist entirely of whole
  genome shotgun sequence reads. Both end sequences and whole genome
  shotgun sequence only contigs will be indicated in the feature
  table.
  ----- Genome Center
  Center: Baylor College of Medicine
  Center code: BCM
  Web site: http://www.hgsc.bcm.tmc.edu/
  Contact: hgsc-help@bcm.tmc.edu
  ----- Project Information
  Center project name: GWB
  Center clone name: CH230-130U10
  ----- Summary Statistics
  Assembly program: Phrap; version 0.990329
  Consensus quality: 222023 bases at least Q40
  Consensus quality: 225500 bases at least Q30
  Consensus quality: 227619 bases at least Q20
  Estimated insert size: 230972; sum-of-contigs estimation
  Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
  -----
  * NOTE: Estimated insert size may differ from sequence length
  * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 5 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 72582: contig of 72582 bp in length
  * 72583 72582: gap of unknown length
  * 72683 196271: contig of 123589 bp in length
  * 196272 196371: gap of unknown length
  * 196372 237682: contig of 41311 bp in length
  * 237683 237782: gap of unknown length
  * 237783 239329: contig of 1547 bp in length

```

```

FEATURES
* 239330 239429: gap of unknown length
* 239430 240477: contig of 1048 bp in length.
Location/Qualifiers
1.240477
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-130J10"
1.1732
/misc_feature
/notes="wgs contig"
1783.3142
/misc_feature
/notes="wgs contig"
72583.72582
gap
/estimated_length=unknown
196272.196371
gap
/estimated_length=unknown
196372.196301
misc_feature
/notes="wgs contig"
237683.237782
gap
/estimated_length=unknown
239330.239429
gap
/estimated_length=unknown

ORIGIN
Query Match 87.2%; Score 21.8; DB 14; Length 240477;
Best Local Similarity 92.0%; Pred. No. 70;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 17869 AAAAAAAAAAGATGCTGTGACGC 17893

RESULT 15
CR405685 116349 bp DNA linear VRT 05-AUG-2005
LOCUS Zebrafish DNA sequence from clone DKEX-164B12 in linkage group 25,
DEFINITION complete sequence.
ACCESSION CR405685
VERSION CR405685.8 GI:71891530
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 116349)
Howden P.
Direct Submission
Submitted (05-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
fish-help@sanger.ac.uk
http://www.sanger.ac.uk/Projects/D_rerio/fage.shtml#dataeighc
On Aug 5, 2005 this sequence version replaced gi:71794019.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
-----

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; information
http://www.sanger.ac.uk/Projects/C_elgans/wormpep
Clone-derived Zebrafish pUC subclones occasionally display

```

```

FEATURES
source
Location/Qualifiers
1.116349
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="25"
/clone="DKEX-164B12"
/clone_11b="DanioKey"

ORIGIN
Query Match 84.0%; Score 21; DB 5; Length 116349;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTG 21
Db 14457 AAAAAAAAAAGCATGCTGTG 14437

RESULT 16
AC002127 144165 bp DNA linear PRI 08-OCT-2003
LOCUS Homo sapiens BAC clone CTA-305H12 from 7, complete sequence.
DEFINITION AC002127
ACCESSION AC002127
VERSION AC002127.1 GI:2121323
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 144165)
Sulston J.E. and Wilson R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
9847074
2 (bases 1 to 144165)
Kepler D. and Wamsley P.
The sequence of Homo sapiens BAC clone CTA-305H12
Unpublished (2001)
3 (bases 1 to 144165)
Waterston R.
Direct Submission
Submitted (27-MAY-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 144165)
Waterston R.
Direct Submission
Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 144165)
Waterston R.
Direct Submission
Submitted (04-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 144165)
Waterston R.
Direct Submission
Submitted (29-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 144165)

```

AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_RG305H12

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHK7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:

Clone CTA-305H12 is from a release of the human BAC library CTB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of CTA-305H12
actual end is at 144165 of CTA-305H12.

FEATURES

source

```
1. 144165
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="7"
   /map="7"
   /clone="CTA-305H12"
   /clone_1kb="CTB-HS-A"
   /rpt_family="AT_rich"
   /rpt_family="MIR"
   /rpt_family="MER2_type"
   /rpt_family="ERV1"
   /rpt_family="ERV1"
   /rpt_family="ERV1"
   /rpt_family="CT-rich"
   /rpt_family="TTTA)n"
   /rpt_family="MERS3"
   /rpt_family="MIR"
   /rpt_family="MIR"
```

```
repeat_region 6933..7338
   /rpt_family="MaLR"
repeat_region 7541..7637
   /rpt_family="MIR"
repeat_region 8015..8393
   /rpt_family="L1"
repeat_region 8395..8435
   /rpt_family="ERV1"
repeat_region 8436..8508
   /rpt_family="(TA)n"
repeat_region 8539..8568
   /rpt_family="L1"
repeat_region 8576..8633
   /rpt_family="L1"
repeat_region 9054..9175
   /rpt_family="L1"
repeat_region 9182..9479
   /rpt_family="MaLR"
repeat_region 9634..9763
   /rpt_family="L1"
repeat_region 9743..10067
   /rpt_family="L1"
repeat_region 10068..10366
   /rpt_family="Alu"
repeat_region 10367..11172
   /rpt_family="L1"
repeat_region 11173..11322
   /rpt_family="L1"
repeat_region 11323..11587
   /rpt_family="L1"
repeat_region 11588..13229
   /rpt_family="L1"
repeat_region 13229..13505
   /rpt_family="L1"
repeat_region 13504..13627
   /rpt_family="L1"
repeat_region 13791..14011
   /rpt_family="MIR"
repeat_region 14135..14219
   /rpt_family="L2"
repeat_region 14369..14394
   /rpt_family="(TTCA)n"
repeat_region 15257..15590
   /rpt_family="ERV1"
repeat_region 15908..16250
   /rpt_family="MaLR"
repeat_region 16293..16360
   /rpt_family="L2"
repeat_region 16362..16419
   /rpt_family="L1"
repeat_region 16497..16683
   /rpt_family="L1"
repeat_region 17097..17183
   /rpt_family="L1"
repeat_region 17195..17554
   /rpt_family="L1"
repeat_region 17787..17806
   /rpt_family="(TTTA)n"
repeat_region 19136..19333
   /rpt_family="MIR"
repeat_region 20370..21007
   /rpt_family="L1"
repeat_region 21152..21431
   /rpt_family="Alu"
repeat_region 21890..22182
   /rpt_family="Alu"
repeat_region 22799..22866
   /rpt_family="Alu"
repeat_region 23352..23724
   /rpt_family="MIR"
repeat_region 24425..24721
   /rpt_family="L1"
repeat_region 25449..25658
   /rpt_family="Alu"
```



```

repeat_region      /rpt_family="TD4"
12173. .12565
/rpt_family="WTA_Mm"
repeat_region      12890. .12918
/rpt_family="AT_rich"
repeat_region      13258. .13435
/rpt_family="B3"
repeat_region      13711. .13857
/rpt_family="B1_Mus1"
repeat_region      13867. .13910
/rpt_family="(CAAAA)n"
repeat_region      14222. .14258
/rpt_family="(GTTG)n"
repeat_region      14748. .14771
/rpt_family="AT_rich"
repeat_region      complement(15704. .16073)
/rpt_family="WTD"
repeat_region      16572. .16833
/rpt_family="B4A"
repeat_region      16844. .16900
/rpt_family="B4A"
repeat_region      18779. .18802
/rpt_family="AT_rich"
repeat_region      18810. .18843
/rpt_family="(GA)n"
repeat_region      complement(18856. .19005)
/rpt_family="RSINE1"
repeat_region      19049. .19163
/rpt_family="B1_Mus3"
repeat_region      19261. .19339
/rpt_family="RSINE1"
repeat_region      19446. .19491
/rpt_family="(TYTA)n"
repeat_region      19493. .19758
/rpt_family="L1M3"
repeat_region      21215. .21250
/rpt_family="(TTTG)n"
repeat_region      complement(21251. .21376)
/rpt_family="B1_Mus8"
repeat_region      21673. .21759
/rpt_family="CT_rich"
repeat_region      22376. .22514
/rpt_family="B3A"
repeat_region      22523. .22546
/rpt_family="(CA)n"
repeat_region      24859. .25078
/rpt_family="B4A"
repeat_region      25085. .25144
/rpt_family="(TC)n"
repeat_region      25145. .25184
/rpt_family="(CA)n"
repeat_region      27834. .27934
/rpt_family="ID_B1"
repeat_region      31973. .32056
/rpt_family="WIRb"
repeat_region      36117. .36151
/rpt_family="(CACG)n"
repeat_region      38825. .38845
/rpt_family="(CAAAA)n"
repeat_region      39695. .39743
/rpt_family="(TG)n"
repeat_region      complement(39996. .40162)
/rpt_family="URRIA"
repeat_region      complement(40275. .41134)
/rpt_family="Lx8"
repeat_region      41135. .41169
/rpt_family="(TG)n"
repeat_region      complement(41170. .41314)
/rpt_family="Lx8"
repeat_region      41462. .41868
/rpt_family="MER68"
repeat_region      42655. .42714
/rpt_family="(TG)n"

```

```

repeat_region      complement(42784. .43066)
/rpt_family="B4A"
repeat_region      46652. .46675
/rpt_family="(GA)n"
repeat_region      46682. .46739
/rpt_family="(TATG)n"
repeat_region      complement(46805. .47066)

Query Match      84.0%; Score 21; DB 9; Length 168310;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAAGCATGACTGTG 21
Db      57854 AAAAAAAAAAGCATGACTGTG 57834

RESULT 18
AC146002      169542 bp      DNA      linear      PRI 18-DEC-2003
LOCUS      Pan troglodytes BAC clone RP43-153B17 from 7, complete sequence.
DEFINITION      AC146002
ACCESSION      AC146002
VERSION      AC146002.2 GI:38016118
KEYWORDS      HTG.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
1 (bases 1 to 169542)
Trani,L., Haglund,K. and Haakenson,W.
The sequence of Pan troglodytes BAC clone RP43-153B17
Unpublished (2001)
2 (bases 1 to 169542)
Wilson,R.
Sequencing of Pan troglodytes
Unpublished (2001)
3 (bases 1 to 169542)
Wilson,R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 169542)
Wilson,R.K.
Direct Submission
Submitted (29-OCT-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 169542)
Wilson,R.
Direct Submission
Submitted (18-DEC-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 29, 2003 this sequence version replaced gi:33386953.

-----Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
-----Summary Statistics
Center project name: C_PT153B17
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

gap /note="assembly_name:Contig14"
13060. .13159
/estimated_length=unknown
misc_feature 13160. .16480
/note="assembly_name:Contig15"
gap 16481. .16580
/estimated_length=unknown
misc_feature 16581. .20758
/note="assembly_name:Contig16"
gap 20759. .20858
/estimated_length=unknown
misc_feature 20859. .25237
/note="assembly_name:Contig17"
gap 25238. .25337
/estimated_length=unknown
misc_feature 25338. .30323
/note="assembly_name:Contig18"
gap 30324. .30423
/estimated_length=unknown
misc_feature 30424. .37205
/note="assembly_name:Contig19"
gap 37206. .37305
/estimated_length=unknown
misc_feature 37306. .44859
/note="assembly_name:Contig20"
gap 44860. .44959
/estimated_length=unknown
misc_feature 44960. .52975
/note="assembly_name:Contig21"
gap 52976. .53075
/estimated_length=unknown
misc_feature 53076. .61583
/note="assembly_name:Contig22"
gap 61584. .61683
/estimated_length=unknown
misc_feature 61684. .72540
/note="assembly_name:Contig23"
gap 72941. .73040
/estimated_length=unknown
misc_feature 73041. .86521
/note="assembly_name:Contig24"
gap 86922. .87021
/estimated_length=unknown
misc_feature 87022. .112076
/note="assembly_name:Contig25"
gap 112077. .112176
/estimated_length=unknown
misc_feature 112177. .139279
/note="assembly_name:Contig26"
gap 139280. .139379
/estimated_length=unknown
misc_feature 139380. .170461
/note="assembly_name:Contig27"
gap 170462. .170561
/estimated_length=unknown
misc_feature 170562. .171882
/note="assembly_name:Contig5"
gap 171883. .171982
/estimated_length=unknown
misc_feature 171983. .173519
/note="assembly_name:Contig9"

ORIGIN
Query Match 84.0%; Score 21; DB 14; Length 173519;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
|||||
Db 44094 AAAAAAAAAAGCATGACTGTG 44074
|||||

RESULT 20

AC005243
LOCUS 175609 bp DNA linear PRI 30-SEP-1998
DEFINITION Homo sapiens chromosome 17, clone hRPK.1124_B_17, complete
sequence.
ACCESSION AC005243
VERSION AC005243.1 GI:3687289
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 175609)
AUTHORS Birren,B., Linton,L., Nuebaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.1124_B_17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175609)
AUTHORS Birren,B., Fasmann,K., Linton,L., Nuebaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatlin,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Gerrigery,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Kann,L.,
Macdonald,P., Marquis,N., McEwan,P., McGuk,A., McKernan,K.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nachman,A., Nafie,R., Naylor,J., O'Connor,T., Pavlin,B.,
Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Strange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Strickland,C., Subramanian,A., Torturella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 175609)
AUTHORS Birren,B., Linton,L., Nuebaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K.,
Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Gerrigery,K., Grant,G., Hagos,B., Healdorf,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehocckj,J., Macdonald,P., Marquis,N., McEwan,P.,
McGuk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nafie,R., Naylor,J., Nloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Strange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Testaye,S., Tichovolsky,N., Torturella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Oct 1, 1998 this sequence version replaced gi:359498.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source
1. .175609
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hRPK.1124_B_17"
/clone_lib="RPCT-11 human BAC library"
/complement(947..1020)
/rpt_family="MIR"
1412..1591
/rpt_family="LTR33"
complement(1592..2109)
/rpt_family="WLR2B"

```

repeat_region complement(2124, .2428)
/rpt_family="Alusx"
repeat_region 2522, .2613
/rpt_family="LTR33"
repeat_region 3063, .3418
/rpt_family="MLT1A2"
repeat_region complement(3587, .3823)
/rpt_family="T199er2"
repeat_region 3825, .3856
/rpt_family="L1P4"
repeat_region 3869, .4191
/rpt_family="MLT2FA"
repeat_region complement(4192, .4304)
/rpt_family="TAGA)n"
repeat_region 4305, .4348
/rpt_family="MLT2FA"
repeat_region complement(6190, .6284)
/rpt_family="MIR"
repeat_region complement(7168, .7374)
/rpt_family="MIR"
repeat_region complement(9857, .10154)
/rpt_family="Alusq"
repeat_region complement(10343, .10509)
/rpt_family="MLT1B"
repeat_region complement(10511, .10604)
/rpt_family="A-rich"
repeat_region complement(10632, .10745)
/rpt_family="L1MA5A"
repeat_region complement(10746, .10852)
/rpt_family="MLT1B"
repeat_region complement(11119, .11223)
/rpt_family="MIR"
repeat_region complement(11858, .12147)
/rpt_family="Aluub"
repeat_region 12149, .12719
/rpt_family="L1MC4"
repeat_region 12724, .12814
/rpt_family="Aluub"
repeat_region 12815, .12847
/rpt_family="CA)n"
repeat_region 12848, .13008
/rpt_family="Aluub"
repeat_region 13009, .13102
/rpt_family="L1MC4"
repeat_region 13274, .13413
/rpt_family="L2"
repeat_region 13471, .13803
/rpt_family="MLT1I"
repeat_region 13804, .14160
/rpt_family="THB1C"
repeat_region 14161, .14206
/rpt_family="MLT1I"
repeat_region 14214, .14313
/rpt_family="L2"
repeat_region complement(14366, .14468)
/rpt_family="L2"
repeat_region 14492, .14599
/rpt_family="L2"
repeat_region 17431, .17473
/rpt_family="MIR"
repeat_region complement(17579, .17662)
/rpt_family="CA)n"
repeat_region 18299, .18364
/rpt_family="L2"
repeat_region 18806, .18859
/rpt_family="MIR"
repeat_region complement(18865, .19057)
/rpt_family="MIR3"
repeat_region 19075, .19233
/rpt_family="MIR"
repeat_region complement(20531, .20730)
/rpt_family="MIR58"
repeat_region complement(20775, .20848)

```

```

repeat_region /rpt_family="L2"
repeat_region complement(21056, .21182)
/rpt_family="MIR3"
repeat_region 21183, .21412
/rpt_family="purine-rich"
repeat_region complement(21485, .21644)
/rpt_family="L2"
repeat_region complement(21715, .21807)
/rpt_family="MIR"
repeat_region complement(21973, .22162)
/rpt_family="MIR3"
repeat_region 22635, .22686
/rpt_family="G-rich"
repeat_region complement(23025, .23078)
/rpt_family="CA)n"
repeat_region 23079, .23113
/rpt_family="CA)n"
repeat_region 24362, .24485
/rpt_family="L2"
repeat_region 24551, .24646
/rpt_family="MIR"
repeat_region 24811, .25177
/rpt_family="L2"
repeat_region 25344, .25532
/rpt_family="L2"
repeat_region 25558, .25797
/rpt_family="L2"
repeat_region 25971, .26230
/rpt_family="AluJo"
repeat_region 26234, .26253
/rpt_family="CAAAA)n"
repeat_region 26296, .26509
/rpt_family="L2"
repeat_region complement(26818, .27125)
/rpt_family="Alusx"
repeat_region complement(27560, .27865)
/rpt_family="AluY"
repeat_region 28416, .28526
/rpt_family="TA)n"
repeat_region complement(28527, .28815)
/rpt_family="AluY"
repeat_region complement(29131, .29578)
/rpt_family="LTR32"
repeat_region complement(29775, .29956)
/rpt_family="7SLRNA"
repeat_region 29965, .30284
/rpt_family="L1PB2"
repeat_region 30505, .30568
/rpt_family="MIR58A"
repeat_region 30569, .30653
/rpt_family="TA)n"
repeat_region 30635, .30793
/rpt_family="MIR58A"
repeat_region 31021, .31059
/rpt_family="AT-rich"

```

```

Query Match 84.0%; Score 21; DB 8; Length 175609;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGCATGACTGTGCA 22
DB 128066 AAAAAAAAAAGCATGACTGTGCA 128086

```

```

RESULT 21
AL954645 AL954645 182141 bp DNA 1linear VRT 12-APR-2003
DEFINITION Zebrafish DNA sequence from clone CH211-23484, complete sequence.
ACCESSION AL954645
VERSION AL954645.7 GI:29824766
KEYWORDS HTG
SOURCE Danio rerio (zebrafish)

```

ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 182141)
 AUTHORS Pelan,S.
 JOURNAL Direct Submission
 Submitted (12-APR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
 On Apr 13, 2003 this sequence version replaced gi:29786467.
 COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormep Repeat names beginning 'dr' were identified by the Recon repeat discovery system (Zhong Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rerio/fishmask.shtml
 CH211-234B4 is from a CHORI-211 BAC library
 VECTOR: PTARBAC2.1.
 FEATURES Location/Qualifiers
 source 1..182141
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-234B4"
 /clone_1ib="CHORI-211"
 ORIGIN
 Query Match 84.0%; Score 21; DB 5; Length 182141;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Y 1 AAAAAAAAAAGCATGCTGTG 21
 ||||||||||||||||||||
 Db 134427 AAAAAAAAAAGCATGCTGTG 134447
 RESULT 22
 LOCUS AC161474 191529 bp DNA linear PRI 22-JUL-2005
 DEFINITION Pan troglodytes BAC clone CH251-488B17 from chromosome 7, complete sequence.
 ACCESSION AC161474
 VERSION AC161474
 KEYWORDS HMG.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Pan.

REFERENCE 1 (bases 1 to 191529)
 AUTHORS Walligorski,J. and Cotton,M.
 JOURNAL The sequence of Pan troglodytes BAC clone CH251-488B17 Unpublished (2001)
 REFERENCE 2 (bases 1 to 191529)
 AUTHORS Wilson,R.K.
 JOURNAL Direct Submission
 Submitted (14-MAY-2005) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 191529)
 AUTHORS Wilson,R.K.
 JOURNAL Direct Submission
 Submitted (03-JUN-2005) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 191529)
 AUTHORS Wilson,R.K.
 JOURNAL Direct Submission
 Submitted (22-JUL-2005) Washington University School of Medicine, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Jul 22, 2005 this sequence version replaced gi:66912508.
 COMMENT ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: C_AB0488B17

 NOTICE:
 This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.
 MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu
 SOURCE INFORMATION:
 The CHORI-251 Chimpanzee BAC library has been constructed at the Children's Hospital Oakland Research Institute, BACPAC Resources, by Dr. Baoji Zhu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Cint', Yerkes #C0471; birthdate:6-6-80). The clone and detailed information can be obtained from Pieter de Jong and co-workers at http://www.bacpac.chori.org.
 ORIGIN
 Query Match 84.0%; Score 21; DB 8; Length 191529;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Y 1 AAAAAAAAAAGCATGCTGTG 21
 ||||||||||||||||||||
 Db 173113 AAAAAAAAAAGCATGCTGTG 173133
 FEATURES Location/Qualifiers
 source 1..191529
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="7"
 /clone="CH251-488B17"
 /clone_1ib="CHORI251"
 This sequence is the entire insert of the clone.

RESULT 23
BV018408
LOCUS
DEFINITION
52126120PFB.T0 CZECHII/E1 Mus musculus STS genomic, sequence
tagged site.
Accession
BV018408
Version
BV018408.1
GI:31102303
Keywords
STS.
Source
Mus musculus (house mouse)
Organism
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
Reference
1 (bases 1 to 819)
Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,
Lander, E.S., Lindblad-Toh, K. and Daly, M.J.
The mosaic structure of variation in the laboratory mouse genome
Nature 420 (6915), 574-578 (2002)
12466852
Title
JOURNAL
PUBMED
COMMENT
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersti@genome.wi.mit.edu
Primer A: None
Primer B: None
STS Size: 819
Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129S1/SVIMJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the MGSCV3 C57BL/6J assembly and SNP
detection was carried out by SSNA-SNP. 225,000 reads were
annotated
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.

FEATURES
source
1. 819
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/E1"
/db_xref="taxon:10090"
/map="10 22-736 80561980-80561266"
/clone_1lb="CZECHII/E1"
<1..>819

ORIGIN
STS
Query Match 83.2%; Score 20.8; DB 10; Length 819;
Best Local Similarity 91.7%; Pred. No. 4.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 781 AAAAAAAAAAGCATGACTGTGACA 804

RESULT 24
AC138524_5
WPCOMMENT
Sequence split into 6 fragments
Fragment Name Begin End
AC138524_0 1 110000
AC138524_1 100001 210000
AC138524_2 200001 310000
AC138524_3 300001 410000
AC138524_4 400001 510000
AC138524_5 500001 550277
Continuation 76 of 6 of AC138524 from base 500001 (AC138524 Homo sapiens chromosome 5

Query Match 83.2%; Score 20.8; DB 14; Length 50277;
Best Local Similarity 91.7%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 2297 AAAAAAAAAAGCATGACTGTGACA 2320

RESULT 25
AC138524_4
WPCOMMENT
Sequence split into 6 fragments
Fragment Name Begin End
AC138524_0 1 110000
AC138524_1 100001 210000
AC138524_2 200001 310000
AC138524_3 300001 410000
AC138524_4 400001 510000
AC138524_5 500001 550277
Continuation 75 of 6 of AC138524 from base 400001 (AC138524 Homo sapiens chromosome 5

Query Match 83.2%; Score 20.8; DB 14; Length 110000;
Best Local Similarity 91.7%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 102297 AAAAAAAAAAGCATGACTGTGACA 102320

RESULT 26
BX908758_1
WPCOMMENT
Sequence split into 6 fragments
Fragment Name Begin End
BX908758_0 1 110000
BX908758_1 100001 210000
BX908758_2 200001 310000
BX908758_3 300001 410000
BX908758_4 400001 510000
BX908758_5 500001 579882
Continuation 72 of 6 of BX908758 from base 100001 (BX908758 Danio rerio clone DKFZ-3202

Query Match 83.2%; Score 20.8; DB 14; Length 110000;
Best Local Similarity 91.7%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 33363 AAAAAAAAAAGCATGACTGTGACA 33386

RESULT 27
AP008210_312/c
WPCOMMENT
Sequence split into 355 fragments
Fragment Name Begin End
AP008210_000 1 110000
AP008210_001 100001 210000
AP008210_002 200001 310000
AP008210_003 300001 410000
AP008210_004 400001 510000
AP008210_005 500001 610000
AP008210_006 600001 710000
AP008210_007 700001 810000
AP008210_008 800001 910000
AP008210_009 900001 1010000
AP008210_010 1000001 1110000
AP008210_011 1100001 1210000
AP008210_012 1200001 1310000
AP008210_013 1300001 1410000

AP008210_014	1400001	1510000	AP008210_087	8700001	8810000
AP008210_015	1500001	1610000	AP008210_088	8800001	8910000
AP008210_016	1600001	1710000	AP008210_089	8900001	9010000
AP008210_017	1700001	1810000	AP008210_090	9000001	9110000
AP008210_018	1800001	1910000	AP008210_091	9100001	9210000
AP008210_019	1900001	2010000	AP008210_092	9200001	9310000
AP008210_020	2000001	2110000	AP008210_093	9300001	9410000
AP008210_021	2100001	2210000	AP008210_094	9400001	9510000
AP008210_022	2200001	2310000	AP008210_095	9500001	9610000
AP008210_023	2300001	2410000	AP008210_096	9600001	9710000
AP008210_024	2400001	2510000	AP008210_097	9700001	9810000
AP008210_025	2500001	2610000	AP008210_098	9800001	9910000
AP008210_026	2600001	2710000	AP008210_099	9900001	10010000
AP008210_027	2700001	2810000	AP008210_100	10000001	10110000
AP008210_028	2800001	2910000	AP008210_101	10100001	10210000
AP008210_029	2900001	3010000	AP008210_102	10200001	10310000
AP008210_030	3000001	3110000	AP008210_103	10300001	10410000
AP008210_031	3100001	3210000	AP008210_104	10400001	10510000
AP008210_032	3200001	3310000	AP008210_105	10500001	10610000
AP008210_033	3300001	3410000	AP008210_106	10600001	10710000
AP008210_034	3400001	3510000	AP008210_107	10700001	10810000
AP008210_035	3500001	3610000	AP008210_108	10800001	10910000
AP008210_036	3600001	3710000	AP008210_109	10900001	11010000
AP008210_037	3700001	3810000	AP008210_110	11000001	11110000
AP008210_038	3800001	3910000	AP008210_111	11100001	11210000
AP008210_039	3900001	4010000	AP008210_112	11200001	11310000
AP008210_040	4000001	4110000	AP008210_113	11300001	11410000
AP008210_041	4100001	4210000	AP008210_114	11400001	11510000
AP008210_042	4200001	4310000	AP008210_115	11500001	11610000
AP008210_043	4300001	4410000	AP008210_116	11600001	11710000
AP008210_044	4400001	4510000	AP008210_117	11700001	11810000
AP008210_045	4500001	4610000	AP008210_118	11800001	11910000
AP008210_046	4600001	4710000	AP008210_119	11900001	12010000
AP008210_047	4700001	4810000	AP008210_120	12000001	12110000
AP008210_048	4800001	4910000	AP008210_121	12100001	12210000
AP008210_049	4900001	5010000	AP008210_122	12200001	12310000
AP008210_050	5000001	5110000	AP008210_123	12300001	12410000
AP008210_051	5100001	5210000	AP008210_124	12400001	12510000
AP008210_052	5200001	5310000	AP008210_125	12500001	12610000
AP008210_053	5300001	5410000	AP008210_126	12600001	12710000
AP008210_054	5400001	5510000	AP008210_127	12700001	12810000
AP008210_055	5500001	5610000	AP008210_128	12800001	12910000
AP008210_056	5600001	5710000	AP008210_129	12900001	13010000
AP008210_057	5700001	5810000	AP008210_130	13000001	13110000
AP008210_058	5800001	5910000	AP008210_131	13100001	13210000
AP008210_059	5900001	6010000	AP008210_132	13200001	13310000
AP008210_060	6000001	6110000	AP008210_133	13300001	13410000
AP008210_061	6100001	6210000	AP008210_134	13400001	13510000
AP008210_062	6200001	6310000	AP008210_135	13500001	13610000
AP008210_063	6300001	6410000	AP008210_136	13600001	13710000
AP008210_064	6400001	6510000	AP008210_137	13700001	13810000
AP008210_065	6500001	6610000	AP008210_138	13800001	13910000
AP008210_066	6600001	6710000	AP008210_139	13900001	14010000
AP008210_067	6700001	6810000	AP008210_140	14000001	14110000
AP008210_068	6800001	6910000	AP008210_141	14100001	14210000
AP008210_069	6900001	7010000	AP008210_142	14200001	14310000
AP008210_070	7000001	7110000	AP008210_143	14300001	14410000
AP008210_071	7100001	7210000	AP008210_144	14400001	14510000
AP008210_072	7200001	7310000	AP008210_145	14500001	14610000
AP008210_073	7300001	7410000	AP008210_146	14600001	14710000
AP008210_074	7400001	7510000	AP008210_147	14700001	14810000
AP008210_075	7500001	7610000	AP008210_148	14800001	14910000
AP008210_076	7600001	7710000	AP008210_149	14900001	15010000
AP008210_077	7700001	7810000	AP008210_150	15000001	15110000
AP008210_078	7800001	7910000	AP008210_151	15100001	15210000
AP008210_079	7900001	8010000	AP008210_152	15200001	15310000
AP008210_080	8000001	8110000	AP008210_153	15300001	15410000
AP008210_081	8100001	8210000	AP008210_154	15400001	15510000
AP008210_082	8200001	8310000	AP008210_155	15500001	15610000
AP008210_083	8300001	8410000	AP008210_156	15600001	15710000
AP008210_084	8400001	8510000	AP008210_157	15700001	15810000
AP008210_085	8500001	8610000	AP008210_158	15800001	15910000
AP008210_086	8600001	8710000	AP008210_159	15900001	16010000

AP008210_160 16000001 16110000
AP008210_161 16100001 16210000
AP008210_162 16200001 16310000
AP008210_163 16300001 16410000
AP008210_164 16400001 16510000
AP008210_165 16500001 16610000
AP008210_166 16600001 16710000
AP008210_167 16700001 16810000
AP008210_168 16800001 16910000
AP008210_169 16900001 17010000
AP008210_170 17000001 17110000
AP008210_171 17100001 17210000
AP008210_172 17200001 17310000
AP008210_173 17300001 17410000
AP008210_174 17400001 17510000
AP008210_175 17500001 17610000
AP008210_176 17600001 17710000
AP008210_177 17700001 17810000
AP008210_178 17800001 17910000
AP008210_179 17900001 18010000
AP008210_180 18000001 18110000
AP008210_181 18100001 18210000
AP008210_182 18200001 18310000
AP008210_183 18300001 18410000
AP008210_184 18400001 18510000
AP008210_185 18500001 18610000
AP008210_186 18600001 18710000
AP008210_187 18700001 18810000
AP008210_188 18800001 18910000
AP008210_189 18900001 19010000
AP008210_190 19000001 19110000
AP008210_191 19100001 19210000
AP008210_192 19200001 19310000
AP008210_193 19300001 19410000
AP008210_194 19400001 19510000
AP008210_195 19500001 19610000
AP008210_196 19600001 19710000

Query Match Best Local Similarity 91.7%; Score 20.8; DB 15; Length 110000;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCTGACTGTGACA 24
DB 99881 AAAAAAAAAATCATCTGTGACA 99858

RESULT 28
CR391975/c 122652 bp DNA linear VRT 26-AUG-2004
LOCUS Zebrafish DNA sequence from clone CH211-276C20 in linkage group 18,
DEFINITION complete sequence.
ACCESSION CR391975
VERSION CR391975.8 GI:51534234
KEYWORDS HTG
SOURCE Danio rerio (zebrafish)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 122652)
REFERENCE
AUTHORS Phillimore, B.
TITLES Direct Submission
JOURNAL Submitted (25-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Aug 24, 2004 this sequence version replaced gi:51470714.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfsh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
CH211-276C20 is from a CHORI-211 BAC library
VECTOR: pTRABAC2.1.

FEATURES

source
1..122652
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-276C20"
/clone_1ib="CHORI-211"

ORIGIN

Query Match Best Local Similarity 91.7%; Score 20.8; DB 5; Length 122652;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCTGACTGTGACA 24
DB 73348 AAAAAAAAAAGATGACTGTGAAA 73325

RESULT 29
OSJN00152/c 133279 bp DNA linear PLN 16-APR-2005
LOCUS Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa008M17,
DEFINITION complete sequence.
ACCESSION AL662950
VERSION AL662950.2 GI:21741778
KEYWORDS HTG
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhacridae; Oryzaceae; Oryza.

REFERENCE
AUTHORS Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J., Liu, Y., Hu, X., Jia, P., Zhang, Y., Zhao, Q., Ying, K., Yu, S., Tang, Y., Weng, Q., Zhang, L., Lu, Y., Mu, J., Lu, Y., Zhang, L., S., Yu, Z., Fan, D., Liu, X., Lu, T., Li, C., Wu, Y., Sun, T., Lei, H., Li, T., Hu, H., Guan, J., Wu, M., Zhang, R., Zhou, B., Chen, Z., Chen, L., Jin, Z., Wang, R., Yin, H., Cai, Z., Ren, S., Lv, G., Gu, W., Zhu, G., Tu, Y., Jia, J., Zhang, Y., Chen, J., Kang, H., Chen, K., Shao, C., Sun, Y., Hu, Q., Zhang, X., Zhang, W., Wang, L., Ding, C., Sheng, H., Gu, J., Chen, S., Ni, L., Zhu, F., Chen, W., Lan, L., Lai, Y., Cheng, Z., Gu, M., Jiang, D., Li, J., Hong, G., Xue, Y. and Han, B.
TITLE Sequence and analysis of rice chromosome 4

JOURNAL
PUBMED
REFERENCE
AUTHORS

Nature 420 (6913), 316-320 (2002)
12447439

2

Han, B., Feng, Q., Huang, Y.C., Li, Y., Zhu, J.J., Zhao, Q., Hu, X.,
Lu, Y.L., Mu, J., Yu, Z., Chen, L., Fan, D.L., Weng, Q.J., Zhang, L.,
Lu, Y.Q., Yu, S.L., Liu, X.H., Lu, T.T., Zhang, Y.J., Lu, Y., Li, C.,
Li, T.T., Zhang, Y., Hu, H., Jia, P.X., Qian, Y.M., Ying, K., Zhou, B.,
Chen, Z.H., Hao, P., Zhang, L., Wu, M., Zhang, R.Q., Guan, J.P., Fu, G.,
Wang, S.Y., Ren, S.X., Lv, G., Lin, W., Gu, W.Q., Zhu, G.F., Tu, X.F.,
Jia, J., Yin, H.F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X.Y.,
Shao, C.Y., Sun, Y., Hu, Q.P., Zhang, X.L., Zhang, W., Wang, L.J.,
Ding, C.W., Sheng, H.H., Gu, J.L., Chen, S.T., Ni, L., Zhu, F.H. and
Hong, G.F.

Direct Submission
Submitted (27-DEC-2001) Han Bin, National Center for Gene Research,
Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
bhan@ncgr.ac.cn

Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
clone: OSJNBa0008M17.
On Jul 12, 2002 this sequence version replaced gi:17998462.
Web site: http://www.ncgr.ac.cn

----- Summary Statistics -----
Assembly program: phrap

This is a complete sequence.
Genes were identified by a combination of several methods: Gene
prediction programs including Fgenesh (<http://www.softberry.com/>),
genSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHMM
(<http://genemark.biology.gatech.edu/Genemark/>), tRNAscan-SE (Sean
Bddy' <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the
complete sequence against NCBI none redundant protein database (nr)
(<ftp://nbi.nlm.nih.gov/biase/db>) and the EST database at NCGR.
Location:Qualifiers

1. 133279

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="japonica"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="4"
/clone="OSJNBa0008M17"
/clone_1id="CG1-OSJNBa"
2503. 6615
/gene="OSJNBa0008M17.1"
join(2503. 2846,3316. 3459,3568. 3651,3793. 3833.
3961. 4005,4323. 4400,4837. 4926,5409. 5563,6370. 6615)
/gene="OSJNBa0008M17.1"
/codon_start=1
/protein_id="CAD11786.2"
/db_xref="GI:38344249"
/translation="MAETKYLSTPRLERLIRGMKGWALIGVACRTLPFRPSLPNQ
RTILPNAQTLLFLMHRCPADVVAENETGALMSDHRKRICLARTGTSSTAIAR
RIITQTEVLSDKIRESMDGVRVAVRRLVADTELAGFLDLPGCWAPHNAVPI
GFGFRSRHVRQORPRPAAVYIAQLYLEMRKCVSSGQPDGLSLGSLDVARMMWAI
LDSLSRAGCESSEHHCWRYPYLRSEKGVADSIYLTQTFGNMTSPKENVYFSSLTVAG
TYMSSQEDCHTENYCOGFIPIRKHOREELAPEDIDPKRQHTLLTREGAALSPSGTL
HANSDELHAASELFWSCGLDEBDKTVDELANAVGTFRAPFILLILLLSQRLRPS
PKDIYIAQIYLSFSA"

gene
CDS
16402. 18911
/gene="OSJNBa0008M17.2"
join(16402. 16553,17149. 17638,18064. 18147,18212. 18283,
18568. 18627,18900. 18911)
/gene="OSJNBa0008M17.2"
/codon_start=1
/protein_id="CAD11787.1"
/db_xref="GI:21741780"
/translation="MDAEGSSDAEAGAGSGKMLRSLDYPCATRLRHRLLAFLINND
YMAVDEINKRRRLVPHVSDLEDMKAGRLVEARVWPRAPPGGGGSGSSAAEAVT
LMKFIHQMLVLDSEAHGIDIDHTAIRGMFRIILAEPPGFSVLNPLVPRPRRLRRRPR

gene
CDS
25005. 30105
/gene="OSJNBa0008M17.3"
join(25005. 25135,25262. 25346,25514. 25586,25691. 25788.
25957. 26049,26682. 26755,27080. 27118,27281. 27269,
27460. 27536,27617. 27691,28039. 28086,28172. 28255,
28335. 28422,28497. 28575,28673. 28763,29325. 29403,
29820. 29933,30031. 30105)
/gene="OSJNBa0008M17.3"
/codon_start=1
/protein_id="CAE04332.2"
/db_xref="GI:38344250"
/db_xref="GOA:Q7XN12"
/db_xref="InterPro:IPR005814"
/translation="MNLIKHAAPAAASFOSETDCTSHASARKFSTSGSSPLDSTENG
FKGSHMLAPPTAAMHSTDLEPLITERSBGSYVDGKNTKLTDLTLAGLMCTALGSEPR
LVKAATDQNLKLPFYHSPFNWSTAKPPLDLAEELISMETAKMGKVFYFNSGEANDSQ
VKLVWYNNALAGRPNKKKIIAOSQAHGSLTISASLGAJPMMLKFDLPAPVLTDC
PHYWRGLPGEAESEFATRLADLENLILKEGETVAAFIAEIVAGGVIPEPKTYF
EKIOAVLOKVDVFIADVEITGGRGLTGMPSGLVNIKDVLAKALSSAVYDCAAT
LVSPEISDVHSGSNKIGFPAHGFYSGHPVSCAVALEAKITIRENIPAHVQIISR
FQGIKAFAPGSSITIGETRGVGLLATEFPANNSPDPFVMEGVADQIFGECKKRGML
VKRVGDEIAMSPPILMSQREVDGLVSIYGEALKATEERVAELRSKKK"

gene
CDS
32352. 37588
/gene="OSJNBa0008M17.4"
join(32352. 32405,33108. 33238,33326. 33410,33553. 33625,
33755. 33859,33986. 34078,34573. 34646,34859. 34897,
35011. 35099,35184. 35260,35351. 35425,35826. 35873,
35960. 36043,36175. 36262,36363. 36441,36531. 36621,
36714. 36792,37086. 37201,37511. 37588)
/gene="OSJNBa0008M17.4"
/codon_start=1
/protein_id="CAE04333.2"
/db_xref="GI:38344251"
/db_xref="GOA:Q7XN11"
/db_xref="InterPro:IPR005814"
/db_xref="InterPro:IPR005814"
/translation="WVIARGLRSLNASSSSSSQAINLKVTSTGSLQGHNTNCDAST
RHSSVSPQVNSTEENGKRGKGLMAPPAQMGSTVPHPLVIRSGSYYYDDCKKY
LBSLAGMCTALGSEPRLVKATDEQLHLPLFHSWNTTKSLDLAEELISMETAK
EMGKVFYFNSGEANDSQVLVWYNNALAGRPDKKFIARSKSYHSTLISALSGIP
ALHOKEDLPAPVLTDCPHYWRGLPGEAESEFATRLADLENLILKEGETVAAFIAE
IAPVMGAGGVIPEPKTYFVKVOAIVKYOILFIADVEITGGRGLTGMPSGLVNIKD
LVSMARALSSAVYDCAATLVSPEISDVHSGSNKIGFPAHGFYSGHPVSCAVALEAK
KITIENITDHPKQIISPRQEGYKAPAGSPIGELTIGVGLIATETPADPNKSPNDPPA
EMGVGAIFGAECQKRGMLVAVADNIMSPPLIMTDEVEELVSIYGDALKATEERVA
ELRSKKN"

gene
CDS
44157. 51427
/gene="OSJNBa0008M17.5"
join(44157. 44234,44839. 45567,45689. 49267,49326. 49616,
49699. 49857,49995. 50153,50457. 50615,50973. 51092,
51206. 51427)
/gene="OSJNBa0008M17.5"
/codon_start=1
/protein_id="CAE04334.2"
/db_xref="GI:38344252"
/db_xref="InterPro:IPR003900"
/db_xref="UniProt/TREMBL:Q7XN10"
/translation="MEVQDAEVEGVTAQKLPDKAVDKAQVDDDKENLVNGSNLQV
KEAHNEDDGTGSDGFELIDVKNPDSAVVEBEKASPTEQGTALAEQALTRBSAM
LNQAEQGLBELFRIIELELKEKLVKYGVEAKNKNTLHYTSQEQRSIADQKEIA
DITQSLKEIYSELSESKERIIOIBELAEIASADKIKHLELKDBERSHAHQARASBL
DITLFTAOIANKMEKPAQIISLOEETKGDQKIDHQVEESIRSTSELKAMRLAKET
LKLENELSAHEBLQAKILNDQMEYKLEQAKKQTPATYLEKQOQILNLTQILDE
SKGNETLCTGIADLNSKLAERDSLRLQAEEDHAKQILDELSKHLDEVLNLSIN
BOHGESRAAENASQKILBELBALQTLHATEERLQKLEAEARVAVAKGSDLEOL
LDESNEKULVASSGELKLEERVOQEAASAEEKLELAEVNTNSVAEKIKLEIQAOLD
STTSKQILQEQVKLSDKFTETHOEPAHSHVHRSLELESLHTSKSDADAVHRTQDL
ENELNANKFEVEVDLEQVRSKYSQSLDELEAYOTKASILEAVNESASEKEKLEVE
SICQITTEKKKLELVLVEYETKEETBEYLEKQSLERLQSOESKVLALLESVLVMEGEK

```

ESHECTIADINLQSLNKNMWMQLESQAGDHSKRTSLSSRAQHEEELINKS
LNDLHVESKTAASALQRIAELETVOEISABOSIKSLHTEPESKLAEEKSDLE
OLKQATEECNSRLVDELSELEEKRTSLKSLSLAOKEMAEISLSEKLDQVNE
KKEFEELSKKATIKHLEANOVALQOLESARHKEEVSDELALGISETLVXLK
SABEOLHEKGALEHATSKKIDLEALQSLLEBDTEMLQOAGENLQOKTECOESAE
LKAESOLASYOAKATAAEVSVVLEAFETITSTLETITBELKTASNAESAE
OALVSRAMSETOALKEIDLAKLMLRELQOPDSTHAKSEVPFKSAHKTIIEHL
TEVHSGELHATKESKNALEKQHEALETTGKQSEVDKNERLALSESISLTLH
VNEAMKOEINAKLVKDELQEKJSSISEKEVAEVAVHEHTIHLREHSGELQ
SABESASBIENELRELTVAQKEAVVDLKEKLVSTETENBKLGINEALKEGELT
KVAMPDELQOESSTHAKKEBAKLVAHERITSHLTETENBKLGINEALKEGELT
SKHLEALMAAOKEAEKDLKSKLDLLEIIGYVEREALREAAAEETHIKEDAVHK
IKSLERQJLAVTENKVELFTEKENLVIANSKNEBELHONKNEQVLAALAAVAEKE
GSEELHSLKTLUDMIQKEBELSOLSLNSTIATYFALRKLSPQVSVVEEHEE
LKSXVITTEEQKLTNEKTESAKKEGEMIAKLEQMVAKSEKELHISKLRQITLS
ELKVEEIQTMQVTEKEDALTTKQEHANVHEKDALQOOLLEVRLKNDAYHTIA
NOBEQVSVREIKMDAYKFSBDLEAHOHVALEEVQVSLQOOLETETHKHEKEO
VSLREYOWEADDSHVSSELKAROVAALEKQIGALTOLOLQVQYRQKTKERRKTL
ALVNTVSKLTQKLSKSAEMKEKIEHLBOKLQAKDSVESTSRDPSLDSSTLRSKORDR
SLAPETTPNPTQOELREPSGIMARKFPLGVALLSVLIGVFUGKRY"
60919. .69865
/gene="OSJNB80008M17.6"
/join(60919. .61386,67375. .67518,68237. .68477,68815. .69315,
69453. .69865)
/gene="OSJNB80008M17.6"
/codon_start=1
/protein_id="CAD41791.2"
/db_xref="GI:38344253"
/db_xref="GGA:07XTS7"
/db_xref="InterPro:IPR001878"
/db_xref="UniProt/TrEMBL:Q7XTS7"

Query Match      83.2%; Score 20.8; DB 15; Length 133279;
Best Local Similarity 91.7%; Pred. No. 2e-02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCTGACTGTGACA 24
Db 121483 AAAAAAAAAATCATCTGACGACA 121460

RESULT 30
AC109193/c AC109193 144093 bp DNA linear ROD 31-JUL-2004
LOCATION Mus musculus chromosome 5, clone RP24-352122, complete sequence.
ACCESSION AC109193
VERSION AC109193.15 GI:50872577
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 144093)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 5, clone RP24-352122
Unpublished
2 (bases 1 to 144093)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chararo,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,M., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagge,B., Horton,L., Hulme,W., Illie,I., Johnson,R., Jones,C.,
Kamat,A., Karacas,A., Kells,C., Lacroque,K., Lamazares,R.,
Lander,T., Lehotzky,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Menue,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunhthang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Roselli,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strausse,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
Toplam,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 144093)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
Dearlano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgeraid,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Illie,I.,
Johnson,R., Jones,C., Kamat,A., Karacas,A., Kells,C., Lander,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Liu,A., Mabbitt,R.,
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrum,J., Menue,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunhthang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Teefaye,S., Theodore,J.,
Toplam,K., Travers,M., Vassiliev,H., Venkatarman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-JUN-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 144093)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
Dearlano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgeraid,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Illie,I.,
Johnson,R., Jones,C., Kamat,A., Karacas,A., Kells,C., Lander,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Liu,A., Mabbitt,R.,
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrum,J., Menue,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunhthang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Teefaye,S., Theodore,J.,
Toplam,K., Travers,M., Vassiliev,H., Venkatarman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-JUL-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 31, 2004 this sequence version replaced gi:58475321.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: I20189
Center clone name: 352_I_22

```


FEATURES

```

source
    Location/Qualifiers
      1..144093
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="5"
        /map="5"
        /clone="RP24-352122"
        /clone_11b="RPc1-24 Male Mouse BAC"
      1..4
        /note="clone boundary
        clone end:SP6
        site:Tbo1"
      repeat_region
        complement(769..903)
        /rpt_family="Lx8"
      repeat_region
        complement(904..1249)
        /rpt_family="OR1A3"
      repeat_region
        complement(1250..1326)
        /rpt_family="Lx8"
      repeat_region
        1756..1800
        /rpt_family="(CAAAAC)n"
      repeat_region
        complement(1909..2041)
        /rpt_family="B2_Mm2"
      repeat_region
        complement(2791..2862)
        /rpt_family="ID_B1"
      repeat_region
        complement(2833..2893)
        /rpt_family="B1F"
      repeat_region
        complement(2861..2917)
        /rpt_family="B1_MM"
      repeat_region
        complement(2924..3048)
        /rpt_family="ID_B1"
      repeat_region
        complement(3361..3426)
        /rpt_family="MIR"
      repeat_region
        complement(3614..3811)
        /rpt_family="Lx"
      repeat_region
        3855..3900
        /rpt_family="(AGGGGG)n"
      repeat_region
        complement(3906..3972)
        /rpt_family="PB1D9"
      repeat_region
        3973..4143
        /rpt_family="(TTCTC)n"
      repeat_region
        4145..4274
        /rpt_family="CT-rich"
      repeat_region
        complement(4286..4350)
        /rpt_family="tRNA-Ala-GCA"
      repeat_region
        complement(4475..4583)
        /rpt_family="Lx9"
      repeat_region
        complement(4615..5053)
        /rpt_family="Lx9"
      repeat_region
        5086..5166
        /rpt_family="(TATG)n"
      repeat_region
        5417..5541
        /rpt_family="(CA)n"
      repeat_region
        6235..6285
        /rpt_family="(CA)n"
      repeat_region
        6874..7048
        /rpt_family="(CATATA)n"
      repeat_region
        7051..7281
        /rpt_family="(CA)n"
      repeat_region
        complement(7330..7702)
        /rpt_family="OR1D"
      repeat_region
        8056..8106
        /rpt_family="(TG)n"
      repeat_region
        complement(8115..8231)
        /rpt_family="RS1NE1"
      repeat_region
        9190..9341
        /rpt_family="ID_B1"
      repeat_region
        9285..9380
        /rpt_family="PB1D10"
      repeat_region
        9650..9693
        /rpt_family="(TG)n"
      repeat_region
        complement(9712..9846)
        /rpt_family="MTE"

```

```

repeat_region      9873..9957
/rpt_family="MER5A"
repeat_region      10137..10302
/rpt_family="(CA)n"
repeat_region      10321..10472
/rpt_family="(CA)n"
repeat_region      10473..10531
/rpt_family="(TC)n"
repeat_region      10531..10566
/rpt_family="(CA)n"
repeat_region      complement(10998..11108)
/rpt_family="MER20"
repeat_region      11421..11472
/rpt_family="(TG)n"
repeat_region      complement(11475..11692)
/rpt_family="B4A"
repeat_region      12353..12452
/rpt_family="MIR"
repeat_region      complement(12553..12739)
/rpt_family="L1M3"
repeat_region      12849..13246
/rpt_family="RMER19B"
repeat_region      13273..13411
/rpt_family="RMER19B"

```

```

Query Match      83.2%; Score 20.8; DB 9; Length 144093;
Best Local Similarity 91.7%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 AAAAAAAAAAGCATGACTGTGACA 24
Db      30532 AAAAAAAAAAGCATGACTGTGACA 30509

```

```

RESULT 31
CR354586
LOCUS      CR354586      144375 bp      DNA      linear      HTG 21-MAR-2005
DEFINITION      Danio rerio clone CH211-142B20, WORKING DRAFT SEQUENCE, 10
unordered pieces.
ACCESSION      CR354586
VERSION      CR354586.5 GI:61673514
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 144375)
REFERENCE
AUTHORS      McLaren,S.
TITLE      Direct Submission
JOURNAL      Submitted (18-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 2005 this sequence version replaced gi:45598153.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZC142B20
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Consensus quality: 140573 bases at least Q40
Consensus quality: 14166 bases at least Q30
Consensus quality: 141642 bases at least Q20
Insert size: 143475; sum-of-contigs
Insert size: 149958; 0.9% error; agarose-fp
Quality coverage: 8.43x in Q20 bases; sum-of-contigs Quality
coverage: 8.14x in Q20 bases; agarose-fp
-----

```

* NOTE: This is a 'working draft' sequence. It currently

* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	31090: contig of 31090 bp in length
*	31091	gap of 100 bp
*	31191	contig of 2954 bp in length
*	31145	gap of 100 bp
*	34245	contig of 29854 bp in length
*	64099	gap of 100 bp
*	64199	contig of 7451 bp in length
*	71650	gap of 100 bp
*	71750	contig of 10419 bp in length
*	82169	gap of 100 bp
*	82269	contig of 9411 bp in length
*	91680	gap of 100 bp
*	91779	gap of 100 bp
*	91780	contig of 21657 bp in length
*	113437	gap of 100 bp
*	113537	contig of 23110 bp in length
*	136647	gap of 100 bp
*	136747	contig of 2349 bp in length
*	138096	gap of 100 bp
*	139196	contig of 5180 bp in length.

FEATURES	Location/Qualifiers
source	1. .144375

```
misc_feature 1. .31090
               /note="assembly_fragment:01825
fragment chain:1
               clone_end:T7
               vector_side:left"
misc_feature 31191. .34144
               /note="assembly_fragment:00039
fragment chain:1"
misc_feature 34245. .64098
               /note="assembly_fragment:01338
fragment chain:1"
misc_feature 64139. .77649
               /note="assembly_fragment:00091
fragment chain:1"
misc_feature 71750. .82168
               /note="assembly_fragment:00227
fragment chain:1"
misc_feature 82269. .91679
               /note="assembly_fragment:00337
fragment chain:1"
misc_feature 91780. .113436
               /note="assembly_fragment:00489
fragment chain:1"
misc_feature 113537. .136646
               /note="assembly_fragment:00509
fragment chain:1"
misc_feature 136747. .139995
               /note="assembly_fragment:00006"
misc_feature 139196. .144375
               /note="assembly_fragment:00150
               clone_end:SP6
               vector_side:right"
```

```

ORIGIN
Query Match      83.2%;   Score 20.8;   DB 14;   Length 144375;
Best Local Similarity 91.7%;   Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 AAAAAAAAAACATGCTGTGACA 24
      |||||

```

Db 90838 AAAAAAAAAAGCATGAGGTGACA 90861

RESULT 32					
AC093755/c					
LOCUS	AC093755	146673 bp	DNA	linear	PRI 13-MAY-2005
DEFINITION	Homo sapiens BAC clone RP11-55C6 from 4, complete sequence.				

VERSION AC093755.2 GI:15638730
KEYWORDS HTG.

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 146673)	Maupin, R. and Shah, N.	The sequence of Homo sapiens BAC clone RP11-55C6

JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 146673)

AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (10-SEP-2001) Genome Sequencing Center, Washington

UNIVERSITY SCHOOL OF MEDICINE, 4444 FOREST PARK PARKWAY, ST. LOUIS
MO 63108, USA
3 (bases 1 to 146673)

AUTHORS	Waterston, R. H.
TITLE	Direct Submission
JOURNAL	Submitted (18-SEP-2001) Genome Sequencing Center, Washington

UNIVERSITY SCHOOL OF MEDICINE, 4444 FOREST PARK PARKWAY, ST. LOUIS,
MO 63108, USA

4 (bases 1 to 146673)

AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington

REFERENCE
AUTHORS
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 146673)
Wilson, R.K.

TITLE Direct Submission
SUBMITTED (13-MAY-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63110

MO 63108, USA
On Sep 18, 2001 this sequence version replaced gi:15529793
----- Genome Center
Corbett Washington Institute for Genome Research

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.edu

----- Summary Statistics -----
Center project name: H_NH0055C06
Drafting Center: WIRB

 DATED: 11/11/2011
 BY: [REDACTED]
 NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For

additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frenken, E.,

Tateno, M., Caranese, J. J. and de Jong, P. J. (1998) An improved

Approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-55C6; actual end is at base position 146673 of RP11-55C6.

The sequence of AC025849 has been incorporated into AC093755.

FEATURES

source

Location/Qualifiers

1. 146673

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="4"

/clone="RP11-55C6"

/clone_lib="RRC1-11"

81397..81658

/note="CpG_island (%GC=62.2, o/e=0.72, #CpGs=21)"

87166..87393

/note="CpG_island (%GC=60.5, o/e=0.81, #CpGs=17)"

134005..134230

/note="CpG_island (%GC=56.2, o/e=0.89, #CpGs=16)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 146673;
Best Local Similarity 91.7%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAAAGCATGCTGACCA 24

Db 81371 AAAAAAAAAAGCATGCTGACCA 81348

RESULT 33

AC155334/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC155334 153509 bp DNA linear ROD 25-FEB-2005
Mus musculus 6 BAC RP24-420024 (Roswell Park Cancer Institute
(C57BL/6J Male) Mouse BAC Library) complete sequence.
AC155334 AC117759
AC155334.7 GI:60279733
HTG.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 153509)
Muzny, D., Adams, C., Abbat, II, O., Allen, C., Albrooks, S., Archer, P.,
Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,
Beraducci, A., Biswal, K., Blyth, P., Boham, H., Buhay, C., Birch, P.,
Cadoree, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,
Chacko, J., Chantour, M., Chavez, D., Chen, A., Chen, G., Chen, R.,
Cheng, M.-T., Chu, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,
Curry, S., Dai, W., Davila, M. L., Davis, C., Davy-Carroil, L., De
Anda, C., Delgado, O., Denson, S., Deramo, G., Ding, Y., Dinh, K.,
Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, H.,
DiLuna, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,
Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,
Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,
Gaskin, C., Gensch, S., Ghose, S., Gill, R., Gonzalez, D.,
Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haebler, K.,
Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P.,
Hawes, A., Hitchens, M., Hayes, S., Hemphill, L., Hernandez, J.,
Hines, S., Hitchens, M., Hodgson, A., Hogues, M., Hollins, B.,
Howell, L. T., Huix, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,
Jacob, J., Jiang, H., Johnson, B., Johnson, R., Kalatsis, K., Kelly, S.,
Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, F.,
Leal, S., Lee, K., Lee, S., Legall, F. I., Lemon, S., Lewis, L., Li, B.,
Li, Y., Li, Z., Jimmell, M., Liu, W., Liu, Y.-S., Liu, Y., Llyanage, D.,
London, P., Lopez, J., Lorensueta, L., Lozano, R., Luk, T., Madu, R.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

FEATURES

source

misc_feature

1..77532 /note="overlaps bases 102274..179805 of clone AC121317"

repeat_region

1784..1848 /function="clone overlap"

repeat_region

2607..2672 /rpt_family="(TG)n"

repeat_region

2767..2927 /rpt_family="tRNA-Ala-GCA"

repeat_region

2955..3051 /rpt_family="Lx8"

repeat_region

3245..3269 /rpt_family="RMRIC"

repeat_region

complement(4219..4914) /rpt_family="AT-rich"

repeat_region

complement(4219..4914) /rpt_family="LI_Mur3"

```

repeat_region complement(4915..5063)
/rpc_family="B3"
repeat_region complement(5064..5152)
/rpc_family="L1_Mur3"
repeat_region complement(5153..5632)
/rpc_family="RLTR25B"
repeat_region complement(5634..5889)
/rpc_family="RLTR25A"
repeat_region complement(6036..6208)
/rpc_family="RLTR25A"
repeat_region complement(6255..7474)
/rpc_family="L1_Mur3"
repeat_region complement(7475..8124)
/rpc_family="RMR12"
repeat_region complement(8125..8281)
/rpc_family="RMR12"
repeat_region complement(8282..9568)
/rpc_family="RMR19B"
repeat_region complement(9569..11084)
/rpc_family="L1_Mur2"
repeat_region complement(11074..11383)
/rpc_family="L1_Mur2"
repeat_region complement(11397..11605)
/rpc_family="L1_Mur3"
repeat_region complement(12253..12608)
/rpc_family="L1_Mur3"
repeat_region complement(12613..12921)
/rpc_family="Lx6"
repeat_region complement(12917..13380)
/rpc_family="Lx6"
repeat_region complement(13381..13707)
/rpc_family="Lx6"
repeat_region complement(13718..13795)
/rpc_family="Lx6"
repeat_region complement(13795..14826)
/rpc_family="Lx6"
repeat_region complement(14826..14963)
/rpc_family="B1_Mm"
repeat_region complement(15320..16073)
/rpc_family="Lx7"
repeat_region complement(16074..16111)
/rpc_family="AT_rich"
repeat_region complement(16160..16250)
/rpc_family="Lx6"
repeat_region complement(16256..16293)
/rpc_family="Lx6"
repeat_region complement(16293..16826)
/rpc_family="TTTTA)n"
repeat_region complement(16826..16957)
/rpc_family="Lx6"
repeat_region complement(16957..17600)
/rpc_family="Lx6"
repeat_region complement(17641..17701)
/rpc_family="Lx6"
repeat_region complement(17702..18365)
/rpc_family="RMR19B"
repeat_region complement(18366..18541)
/rpc_family="Lx6"
repeat_region complement(18608..18688)
/rpc_family="GA)n"
repeat_region complement(18688..19593)
/rpc_family="CT_rich"
repeat_region complement(19593..21368)
/rpc_family="Lx6"
repeat_region complement(21368..22341)
/rpc_family="Lx6"
repeat_region complement(22341..22795)
/rpc_family="Lx6"
repeat_region complement(22795..22821)
/rpc_family="Lx6"
repeat_region complement(22821..23331)
/rpc_family="Lx6"
repeat_region complement(23331..23349)
/rpc_family="Lx6"
repeat_region complement(23349..24524)
/rpc_family="L1_Mur3"
repeat_region complement(24524..24592)
/rpc_family="L1_Mur3"
repeat_region complement(24592..25142)
/rpc_family="L1_Mur3"

```

```

repeat_region /rpc_family="RMR2"
25266..25399
/rpc_family="L1_MEC"
repeat_region complement(26242..26380)
/rpc_family="RMR1"
26424..26471
/rpc_family="AT_rich"
27049..27667
/rpc_family="RMR17"
27885..27911
/rpc_family="CAAA)n"
repeat_region complement(28734..29132)
/rpc_family="AT_rich"
29144..29166
/rpc_family="AT_rich"
29964..30052
/rpc_family="TG)n"
repeat_region complement(30055..30109)
/rpc_family="B4"

Query Match 83.2% Score 20.8; DB 9; Length 153509;
Best Local Similarity 91.7%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACATGACTGTGACA 24
DB 87538 AAAAAAGAAACATGACTGTGACA 87515

RESULT 34
LOCUS BX005050/166695 bp DNA linear VRT 15-JUL-2003
DEFINITION Zebrafish DNA sequence from clone CH211-11A11, complete sequence.
ACCESSION BX005050
VERSION BX005050.6 GI:32452253
KEYWORDS HTG.
SOURCE Dario reio (zebrafish)
ORGANISM Dario reio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 166695)
Sycamore, N.
Direct Submission
Submitted (15-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 4, 2003 this sequence version replaced gi:32398451.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPP; Information
on the WORMPP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpp/Clone-derived

```

zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirony Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_zebrafish/fishmask.shtml CH211-11A11 is from a CHORI-211 BAC library

FEATURES

source

1. 166695
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-11A11"
/clone_11b="CHORI-211"

ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 166695;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAAGCATGACTGTGAC 25

Db 159038 AAAAAAAAAAGCATGACTGTGAAAC 159015

RESULT 35

CR788236 180541 bp DNA linear HTG 03-DEC-2004
LOCUS CR788236
DEFINITION Danio rerio clone DKEX-92F12, WORKING DRAFT SEQUENCE.
ACCESSION CR788236.5 GI:56310024
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULFILL.
KEYWORDS Danio rerio (zebrafish)
SOURCE Danio rerio
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 180541)
Phillimore, B.
Direct Submission
Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
On Dec 3, 2004 this sequence version replaced gi:55469004.

REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zK92F12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 178654 bases at least Q40
Consensus quality: 179027 bases at least Q30
Consensus quality: 179208 bases at least Q20
Insert size: 180541; sum-of-contigs
Insert size: 178301; 4.8% error; agarose-fp
Quality coverage: 6.26x in Q20 bases; sum-of-contigs Quality
coverage: 6.65x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced

* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 180541: contig of 180541 bp in length.
Location/Qualifiers

FEATURES

source

1. 180541
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-92F12"
/clone_11b="DanioKey"
1. 180541
/note="assembly_fragment:02267"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 180541;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGACTGTGACA 24

Db 97841 AAAAAAAAAAGCATGACTGTGAAA 97864

RESULT 36

AC073253/c 187045 bp DNA linear HTG 16-JUN-2000
LOCUS AC073253
DEFINITION Homo sapiens chromosome 8 clone RP11-15816, WORKING DRAFT SEQUENCE.
ACCESSION AC073253.1 GI:8468969
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 187045)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 187045)
Waterston, R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

----- Genome Center
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1566: contig of 1566 bp in length
* 1567 1666: gap of unknown length
* 1667 3438: contig of 1772 bp in length
* 3439 3538: gap of unknown length
* 3539 5107: contig of 1569 bp in length
* 5108 5207: gap of unknown length
* 5208 8079: contig of 2872 bp in length
* 8080 8179: gap of unknown length
* 8180 12279: contig of 4100 bp in length
* 12280 12379: gap of unknown length
* 12380 16847: contig of 4468 bp in length
* 16848 16947: gap of unknown length
* 16948 28772: contig of 11825 bp in length
* 28773 28872: gap of unknown length
* 28873 41443: contig of 12571 bp in length
* 41444 41543: gap of unknown length
* 41544 60821: contig of 19278 bp in length

```

* 60822 60921: gap of unknown length
* 60922 86187: contig of 25266 bp in length
* 86188 86287: gap of unknown length
* 86288 110231: contig of 23944 bp in length
* 110232 110331: gap of unknown length
* 110332 145139: contig of 34808 bp in length
* 145140 145239: gap of unknown length
* 145240 187045: contig of 41806 bp in length.
Location/Qualifiers
source
1. 187045
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-15816"
1567. 1666
/estimated_length=unknown
3439. 3538
/estimated_length=unknown
5108. 5207
/estimated_length=unknown
8080. 8179
/estimated_length=unknown
12280. 12379
/estimated_length=unknown
16848. 16947
/estimated_length=unknown
28773. 28872
/estimated_length=unknown
41444. 41543
/estimated_length=unknown
60822. 60921
/estimated_length=unknown
86188. 86287
/estimated_length=unknown
110232. 110331
/estimated_length=unknown
145140. 145239
/estimated_length=unknown

```

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 187045;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCAGTGAAGTGTGACA 24

DB 128552 AAAAAAAAAAGCAGTGAAGTGTGACA 128529

RESULT 37
AC137261/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-unknown, WORKING DRAFT SEQUENCE, 4
unordered pieces.

AC137261 187819 bp DNA linear HTG 19-NOV-2002
AC137261
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 187819)

REFERENCE
AUTHORS
Muzny, D., Marle, Metzger, M., Lee, A., Adams, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Bunay, C., Burch, P., Butrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Casar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
2 (bases 1 to 187819)
Rat Genome Sequencing Consortium.
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: K2HR
----- Summary Statistics
Assembly program: Phrap; version 0.950329
Consensus quality: 173722 bases at least Q40
Consensus quality: 175280 bases at least Q30
Consensus quality: 176398 bases at least Q20

Estimated insert size: 175471; sum-of-ctnigs estimation
Quality coverage: 8x in Q20 bases; sum-of-ctnigs estimation

NOT: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 4 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 121089: contig of 121089 bp in length
* 121090 121189: gap of unknown length
* 121190 185280: contig of 64091 bp in length
* 185281 185380: gap of unknown length
* 185381 186528: contig of 1148 bp in length
* 186529 186628: gap of unknown length
* 186629 187819: contig of 1191 bp in length.

FEATURES

Location/Qualifiers

1..187819
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-unknown"

misc_feature

1..1131
/note="wgs contig"

gap

121090..121189
/estimated_length=unknown

misc_feature

121190..122462
/note="wgs contig"

gap

185281..185380
/estimated_length=unknown

gap

186529..186628
/estimated_length=unknown

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 187819;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24

Db 160076 AAAAAAAAAAGCATGACTGTGACA 160053

RESULT 38

EX530070

LOCUS 191146 bp DNA linear VRT 29-JAN-2004
DEFINITION Zebrafish DNA sequence from clone DKEY-245C12 in linkage group 7,
complete sequence.

ACCESSION

EX530070

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (15-JAN-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk
On Jan 17, 2004 this sequence version replaced gi:36304125.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfsh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhifeng Bao and Sean Eddy, submitted), and those
beginning 'dir' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-245C12
is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.

FEATURES

Location/Qualifiers

1..191146
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-245C12"
/clone_lib="DanioKey"

ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 191146;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24

Db 113978 AAAAAAAAAAGCATGACTGTGACA 114001

RESULT 39

AC109022/c

LOCUS 194439 bp DNA linear HTG 11-OCT-2002

DEFINITION Rattus norvegicus clone CH230-273A9, *** SEQUENCING IN PROGRESS
***, 6 unordered pieces.

ACCESSION

AC109022

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Muzny, D., Maric, M., Metzger, M., Lee, S., Abmayr, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amiri, A., Anguiano, D.,
Anyalebech, V., Ayagi, A., Ayodeji, M., Baca, B., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Bunay, C., Butch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy, C., De Ardo, C., Dedrich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunatirane, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A., Jarkup, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowls, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzen, L., Louie, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nait, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokoelam, O., Okunolu, G., Olajunmugbon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Sma's, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J., Steele, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villalana, D., Waldron, J., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wlasczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE

Journal
2 (bases 1 to 194439)

REFERENCE

Worley, K.C.

AUTHORS

Worley, K.C.

JOURNAL

Submitted (03-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 194439)

REFERENCE

Rat Genome Sequencing Consortium.

AUTHORS

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL

On Oct 9, 2002 this sequence version replaced g1:21737747.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GPRD

Center clone name: CH230-273A9

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 154759 bases at least Q40
Consensus quality: 159418 bases at least Q30
Consensus quality: 162164 bases at least Q20
Estimated insert size: 162030; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 77397: contig of 77397 bp in length
* 77398 77497: gap of unknown length
* 77498 11192: contig of 33695 bp in length
* 11193 11292: gap of unknown length
* 11293 189462: contig of 78170 bp in length
* 189463 189562: gap of unknown length
* 189563 190843: contig of 1281 bp in length
* 190844 190943: gap of unknown length
* 190944 192153: contig of 1210 bp in length
* 192154 192254: gap of unknown length
* 192254 194439: contig of 2186 bp in length.

FEATURES
source
1..194439
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-273A9"
341..1146
/note="clone boundary
clone_end:177
site:Mbol
end_sequence:RXAF105TV"
13768..14271
/note="clone boundary
clone_end:Sp6
site:Mbol
end_sequence:RXAF105TV"
77398..77497
/estimated_length=unknown
11193..11292
/estimated_length=unknown
11293..112527
/note="wgs_end_extension
clone_end:Sp6"
189463..189562
/estimated_length=unknown
190844..190943
/estimated_length=unknown
192154..192253
/estimated_length=unknown

misc_feature
gap
gap
misc_feature
gap
gap
gap
ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 194439;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACATGCTGACCA 24
Db 35409 AAAAAAAAAAATCACTGTGACA 35386

RESULT 40
AC161795/c
LOCUS
DEFINITION Mus musculus chromosome 18 clone R23-121G22 map 18, *** SEQUENCING

IN PROGRESS ***, 7 unordered pieces.

ACCESSION
ACT61795 GI:68563551
HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 208833)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 18, clone RP23-121G22
Unpublished
2 (bases 1 to 208833)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArrelano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Etickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatae,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
Maclean,C., MacDonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Punukhang,P., Pierre,N., Rachupka,A., Ramsamy,U., Raymond,C.,
Reita,R., Rise,C., Rogov,P., Roman,S., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Testafaye,S., Theodore,J.,
Tophan,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (20-May-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208833)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,
Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArrelano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Etickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatae,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
Maclean,C., MacDonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Punukhang,P., Pierre,N., Rachupka,A., Ramsamy,U., Raymond,C.,
Reita,R., Rise,C., Rogov,P., Roman,S., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Testafaye,S., Theodore,J.,
Tophan,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (05-JUL-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
On Jul 5, 2005 this sequence version replaced gi:66954872.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: WTB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information

Center project name: L32528
Center clone name: 121_G22

NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 24346: contig of 24346 bp in length
* 24347 24446: gap of unknown length
* 24447 24447: gap of unknown length
* 54070 54169: gap of unknown length
* 54170 71006: contig of 16837 bp in length
* 71007 71106: gap of unknown length
* 71107 89106: contig of 18000 bp in length
* 89107 89206: gap of unknown length
* 89207 91845: contig of 2639 bp in length
* 91846 91845: gap of unknown length
* 91946 188249: contig of 96304 bp in length
* 188250 188349: gap of unknown length
* 188350 208833: contig of 20484 bp in length.

FEATURES
source
1. 208833
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/map="18"
/clone="RP23-121G22"
/clone_lib="RP23-121G22 Female Mouse BAC"
24347..24446
/estimated_length=unknown
54070..54169
/estimated_length=unknown
71007..71106
/estimated_length=unknown
89107..89206
/estimated_length=unknown
91846..91945
/estimated_length=unknown
188250..188349
/estimated_length=unknown

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 208833;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAAGCATGACGTGACCA 24
Db 87188 AAAAAAAAAAGCATGACGTGACCA 87165

RESULT 41
LOCUS
AC161925
DEFINITION
Mus musculus chromosome 18 clone RP23-5915 map 18, WORKING DRAFT
SEQUENCE 35 unordered pieces.
AC161925
VERSION
AC161925.1 GI:66392667
KEYWORDS
HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 213666)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 18, clone RP23-5915
Unpublished

REFERENCE
AUTHORS
2 (bases 1 to 213666)
Birken, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Archach, H.M., Barua, N., Bastien, V.,
Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J.,
Chapel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
DeRellano, K., Diaz, D.S., Dodge, S., Dooley, K., Dorris, J.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hegos, B., Hall, J., Horton, L., Hulme, M., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelis, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Menes, L., Mhova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhthang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Reita, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Stubbs, M., Talmas, J., Teefaye, S., Theodore, J.,
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Vael, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemek, L.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAY-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@broad.mit.edu

----- Project Information
Project name: L32553
Center clone name: 59_L5
Center clone name: 59_L5

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 200447 bases at least Q40
Consensus quality: 205025 bases at least Q30
Consensus quality: 208388 bases at least Q20
Insert size: 230000; agarose-fp
Insert size: 210266; sum-of-coverage
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1333: contig of 1333 bp in length
1334
1433: gap of unknown length
1434
1433: contig of 1018 bp in length
2451: gap of unknown length
2452
2551: gap of unknown length
2552
3569: contig of 1018 bp in length
3569: gap of unknown length
3570
3669: gap of unknown length
3669: contig of 2226 bp in length
5895: gap of unknown length
5895: gap of unknown length
7002: contig of 1007 bp in length
7002: gap of unknown length
7003
7102: gap of unknown length
7102: contig of 3052 bp in length
10154: gap of unknown length
10155
10254: gap of unknown length
10254: contig of 2948 bp in length
13203
13302: gap of unknown length
13302: contig of 2040 bp in length
15342: gap of unknown length
15343
15442: gap of unknown length
15443
18266: contig of 2824 bp in length

18366: gap of unknown length
18367
18367: contig of 2451 bp in length
20818
20817: gap of unknown length
20918
24056: contig of 3139 bp in length
24057
24156: gap of unknown length
24157
27612: contig of 3456 bp in length
27613
27712: gap of unknown length
30773
30777: contig of 3065 bp in length
30778
33607: gap of unknown length
33607: contig of 2730 bp in length
33608
33707: gap of unknown length
33707: contig of 2925 bp in length
36632
36632: gap of unknown length
36633
36732: gap of unknown length
36733
40506: contig of 3774 bp in length
40507
40606: gap of unknown length
43166: contig of 2560 bp in length
43167
43266: gap of unknown length
43267
48355: contig of 5089 bp in length
48356
48355: gap of unknown length
48456
51702: contig of 3247 bp in length
51703
51802: gap of unknown length
51803
57203: contig of 5401 bp in length
57204
57303: gap of unknown length
57304
59979: contig of 2676 bp in length
59980
60079: gap of unknown length
60080
63103: contig of 3024 bp in length
63104
63203: gap of unknown length
63204
67272: contig of 4069 bp in length
67273
67373: gap of unknown length
67373
67452: gap of 80 bp in length
67453
67552: gap of unknown length
67553
72895: contig of 5343 bp in length
72896
72995: gap of unknown length
79469
79469: contig of 6474 bp in length
79470
79569: gap of unknown length
79570
87437: contig of 7868 bp in length
87438
87537: gap of unknown length
87538
95675: contig of 8138 bp in length
95676
95775: gap of unknown length
95776
103896: contig of 8121 bp in length
103897
103966: gap of unknown length
103967
112219: contig of 8223 bp in length
112220
112319: gap of unknown length
112319: gap of unknown length
126518: contig of 1419 bp in length
126519
126518: gap of unknown length
126519
139657: contig of 13039 bp in length
139658
139757: gap of unknown length
139758
155665: contig of 15908 bp in length
155666
155785: gap of unknown length
155786
173925: contig of 18160 bp in length
173926
213666: gap of unknown length
213666: contig of 39641 bp in length.

FEATURES
source
1. 213666
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/map="18"
/clone="RP23-5915"
/clone_1="RP23-5915"
1. 1333
/note="assembly_fragment"
1334. 1433
/estimated_length=unknown
1434. 2451
/note="assembly_fragment"
1434. 2551
/estimated_length=unknown
2452. 2551
/estimated_length=unknown
2552. 3569
/note="assembly_fragment"
3570. 3669
/estimated_length=unknown

misc_feature
1. 1333
/clone="RP23-5915"
/clone_1="RP23-5915"
1334. 1433
/note="assembly_fragment"
1434. 2451
/estimated_length=unknown
2452. 2551
/estimated_length=unknown
2552. 3569
/note="assembly_fragment"
3570. 3669
/estimated_length=unknown

misc_feature
1. 1333
/clone="RP23-5915"
/clone_1="RP23-5915"
1334. 1433
/note="assembly_fragment"
1434. 2451
/estimated_length=unknown
2452. 2551
/estimated_length=unknown
2552. 3569
/note="assembly_fragment"
3570. 3669
/estimated_length=unknown

misc_feature
1. 1333
/clone="RP23-5915"
/clone_1="RP23-5915"
1334. 1433
/note="assembly_fragment"
1434. 2451
/estimated_length=unknown
2452. 2551
/estimated_length=unknown
2552. 3569
/note="assembly_fragment"
3570. 3669
/estimated_length=unknown

```

misc_feature      3670..5895
                   /note="assembly_fragment"
gap               5895..5995
                   /estimated_length=unknown
misc_feature      5995..7002
                   /note="assembly_fragment"
gap               7002..7102
                   /estimated_length=unknown
misc_feature      7102..10154
                   /note="assembly_fragment"
gap               10154..10254
                   /estimated_length=unknown
misc_feature      10254..13702
                   /note="assembly_fragment"
gap               13702..13702
                   /note="assembly_fragment"
misc_feature      13702..13702
                   /estimated_length=unknown
misc_feature      13702..15342
                   /note="assembly_fragment"
gap               15342..15442
                   /estimated_length=unknown
misc_feature      15442..18266
                   /note="assembly_fragment"
gap               18266..18366
                   /estimated_length=unknown
misc_feature      18366..20817
                   /note="assembly_fragment"
gap               20817..20917
                   /estimated_length=unknown
misc_feature      20917..24056
                   /note="assembly_fragment"
gap               24056..24156
                   /estimated_length=unknown
misc_feature      24156..27612
                   /note="assembly_fragment"
gap               27612..27712
                   /estimated_length=unknown
misc_feature      27712..30777
                   /note="assembly_fragment"
gap               30777..30877

```

```

Query Match      83.2%; Score 20.8; DB 14; Length 213666;
Best Local Similarity 91.7%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAGCATGCTGACCA 24
Db      74529 AAAAAAAAAAGCATGATGTGACA 74552

```

```

RESULT 42
LOCUS      AC121411/c      223824 bp      DNA      linear      HTG 15-NOV-2002
DEFINITION      Rattus norvegicus clone CH230-524J18, WORKING DRAFT SEQUENCE.
ACCESSION      AC121411
VERSION      AC121411.4 GI:25008224
KEYWORDS      HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
            1 (bases 1 to 223824)
            Muzny, D., Maric, M., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
            Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,
            Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
            Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benamed, F.,
            Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
            Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
            Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
            Chacón, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
            Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Crege, A., D'Souza, L.,
            Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
            Delgado, O., Denson, S., Detamo, C., Ding, Y., Dinh, H., Diya, K.,

```

```

TITLE      JOURNAL
AUTHORS      REFERENCE
TITLE      JOURNAL
AUTHORS      JOURNAL
TITLE      JOURNAL

```

REFERENCE

AUTHORS

TITLE

COMMENT

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Gabis, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunartne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensunhewa, L., Louisedge, H., Lozada, R., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmood, M., Malloy, K., Mangum, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokediemeh, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeym, A., Sison, I., Sitter, C.D., Smaiz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Swatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., and Gibbs, R.A.

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

```

----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 208022 bases at least Q40
Consensus quality: 209828 bases at least Q30
Consensus quality: 210923 bases at least Q20
Estimated insert size: 212739, sum-of-contigs estimation
Quality coverage: 7x in Q20 bases, sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) )
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*   1  223824: contig of 223824 bp in length.
FEATURES             location/Qualifiers
    source            1..223824
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-524718"
                        1..1086
misc_feature          /note="wgs_end_extension
                      clone_end:Sp6"
                      5407..6697
misc_feature          /note="wgs_end_extension
                      clone_end:Sp6"
                      7242..8131
misc_feature          /note="Clone boundary
                      clone_end:Sp6
                      site:
                      end_sequence:BZ187350"
                      215537..216657
                      /note="Clone boundary
                      clone_end:T7
                      site:
                      end_sequence:BZ187350"
                      222141..223824
                      /note="wgs_end_extension
                      clone_end:T7"
misc_feature          misc_feature
                      misc_feature
ORIGIN
Query Match         83.2%; Score 20.8; DB 14; Length 223824;
Best Local Similarity 91.7%; Pred.No.1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAGCATGCTGTGCACA 24
Dn 102333 AAAAAAAAAAGATTACTGTGCACA 102310

RESULT 43
BX537105/c          BX537105      225581 bp      DNA       linear      HTG 06-AUG-2003
LOCUS              BX537105/c
ACCESSION          BX537105
VERSION            BX537105.2 GI:31559379
KEYWORDS           HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE             Danio rerio (zebrafish)
ORGANISM           Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 225581)
Burton,J.
Direct Submission
Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

```

COMMENT

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jun 9, 2003 this sequence version request replaced gi:11411851.

----- Genome Center -----

Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information -----

Center project name: zKJL24
 ----- Summary Statistics -----

Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 217427 bases at least Q40
 Consensus quality: 219130 bases at least Q30
 Consensus quality: 220818 bases at least Q20
 Insert size: 223881; sum-of-contigs
 Insert size: 219000; 7.3% error; agarose-fp
 Quality coverage: 6.82x in Q20 bases; sum-of-contigs Quality
 coverage: 7.43x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	20602: contig of 20602 bp in length
*	20603
*	20702: gap of 100 bp
*	20703
*	57831: contig of 37129 bp in length
*	57832
*	57931: gap of 100 bp
*	57932
*	65164: contig of 7233 bp in length
*	65165
*	65265
*	71392: contig of 6128 bp in length
*	71393
*	71492: gap of 100 bp
*	71493
*	93677: contig of 22185 bp in length
*	93678
*	93777: gap of 100 bp
*	93778
*	101599: contig of 7821 bp in length
*	101599
*	101698: gap of 100 bp
*	101699
*	128689: contig of 26991 bp in length
*	128690
*	128789: gap of 100 bp
*	128790
*	138632: contig of 9843 bp in length
*	138633
*	138732: gap of 100 bp
*	138733
*	144193: contig of 5461 bp in length
*	144194
*	144293: gap of 100 bp
*	144294
*	155818: contig of 11525 bp in length
*	155819
*	155918: gap of 100 bp
*	155919
*	168354: contig of 12436 bp in length
*	168355
*	168454: gap of 100 bp
*	168455
*	171301: contig of 2847 bp in length
*	171302
*	171401: gap of 100 bp
*	171402
*	180364: contig of 8963 bp in length
*	180365
*	180464: gap of 100 bp
*	180465
*	188604: contig of 8140 bp in length
*	188605
*	188704: gap of 100 bp
*	188705
*	192655: contig of 4151 bp in length
*	192656
*	192955: gap of 100 bp
*	192956
*	201754: contig of 8799 bp in length
*	201755
*	201854: gap of 100 bp
*	201855
*	211372: contig of 9518 bp in length
*	211373
*	211472: gap of 100 bp
*	211473
*	225581: contig of 14109 bp in length.

Location/Qualifiers

1. 225581

/organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone_id="DXEY-3124"
 /clone_lib="Daniokey"
 1. 20602
 /note="assembly_fragment:01463
 fragment chain:1"

FEATURES

source

misc_feature

```

misc_feature      20703..57831
                  /note="assembly_fragment:01892
                  fragment_chain:1"
misc_feature      57932..65164
                  /note="assembly_fragment:02319
                  fragment_chain:1"
misc_feature      65265..71392
                  /note="assembly_fragment:02166
                  fragment_chain:1"
misc_feature      71493..93677
                  /note="assembly_fragment:01041
                  fragment_chain:1"
misc_feature      93778..101598
                  /note="assembly_fragment:02235
                  fragment_chain:1"
misc_feature      101699..128689
                  /note="assembly_fragment:01236
                  fragment_chain:1"
misc_feature      128790..138632
                  /note="assembly_fragment:02758
                  fragment_chain:1"
misc_feature      138733..144193
                  /note="assembly_fragment:02245
                  fragment_chain:1"
misc_feature      144294..155818
                  /note="assembly_fragment:01852
                  fragment_chain:1"
misc_feature      155919..168354
                  /note="assembly_fragment:01722
                  fragment_chain:1"
misc_feature      168455..171301
                  /note="assembly_fragment:00772
                  fragment_chain:2"
misc_feature      171402..180364
                  /note="assembly_fragment:02453
                  fragment_chain:2"
misc_feature      180465..188604
                  /note="assembly_fragment:00524
                  fragment_chain:2"
misc_feature      188705..192855
                  /note="assembly_fragment:02841
                  fragment_chain:2"
misc_feature      192956..201754
                  /note="assembly_fragment:00644
                  fragment_chain:2"
misc_feature      201855..211372
                  /note="assembly_fragment:02536
                  fragment_chain:2"
misc_feature      211473..225581
                  /note="assembly_fragment:01373
                  fragment_chain:2"

ORIGIN

Query Match      83.2% Score 20.8; DB 14; Length 225581;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGACTGTGACA 24
Db      123690 AAAAAAAAAAGCATGAGTGACA 123667

RESULT 44
AC152773/c      227705 bp DNA linear HTG 01-JUL-2005
LOCUS          AC152773
DEFINITION    Bos taurus clone CH240-15N19, *** SEQUENCING IN PROGRESS ***, 28
unordered pieces.
AC152773
VERSION      AC152773.4 GI:68226853
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Bos taurus
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 227705)
Allen, C., Allen, H., Albrooke, S., Amin, A., Anguiano, D.,
Ayalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blythe, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Canter, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Dublin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falla, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisti, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,
Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokeleh, O., Okwum, G., Olamunagbon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Phannoch, C.,
Plopper, F., Polindexter, A., Popovic, D., Primm, E., Pu, L., L.,
Puzo, M., Qutro, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, L., Sitter, C.D., Sma's, D.,
Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Soes, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabori, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemari, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Wortley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 227705)
Worley, K.C.
Direct Submission
Submitted (18-NOV-2004) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 227705)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rac/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated

```

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: PARU
Center clone name: CH240-15N19
----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 215316 bases at least Q40
Consensus quality: 217686 bases at least Q30
Consensus quality: 220035 bases at least Q20
Estimated insert size: 223887; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1
3270      3269: contig of 3269 bp in length
3822      3821: gap of 552 bp
6292      6291: contig of 2470 bp in length
6502      6500: gap of 209 bp
6501      10078: contig of 3578 bp in length
10079      10128: gap of 50 bp
10129      20018: contig of 9890 bp in length
20019      20068: gap of 50 bp
20069      24148: contig of 4080 bp in length
24149      24198: gap of 50 bp
24199      26016: contig of 1818 bp in length
26017      26656: gap of 640 bp
26657      32724: contig of 6068 bp in length
32725      32774: gap of 50 bp
32775      46963: contig of 14189 bp in length
46964      47013: gap of 50 bp
47014      52515: contig of 5502 bp in length
52516      52565: gap of 50 bp
52566      62835: contig of 10270 bp in length
62836      63012: gap of 177 bp
63013      77338: contig of 14326 bp in length
77339      77388: gap of 50 bp
77389      81137: contig of 3749 bp in length
81138      81187: gap of 50 bp
81188      85197: contig of 4010 bp in length
85198      85247: gap of 50 bp
85248      92912: contig of 7665 bp in length
92913      93132: gap of 220 bp
93133      97318: contig of 4186 bp in length
97319      97418: gap of unknown length
97419      106826: contig of 9408 bp in length
106827      11368: contig of 4492 bp in length
11369      11418: gap of 50 bp
11419      113711: contig of 2293 bp in length
113712      113761: gap of 50 bp
113762      115572: contig of 1811 bp in length
115573      115672: gap of unknown length
115681      120681: contig of 5009 bp in length
120682      120731: gap of 50 bp
120732      149048: contig of 28317 bp in length
```

```
* 149049      149802: gap of 754 bp
* 149803      180393: contig of 30591 bp in length
* 180394      180443: gap of 50 bp
* 180444      192889: contig of 12446 bp in length
* 192890      193963: gap of 1074 bp
* 193964      201891: contig of 7928 bp in length
* 201892      201941: gap of 50 bp
* 201942      222969: contig of 21028 bp in length
* 222970      223069: gap of unknown length
* 223070      224438: contig of 1369 bp in length
* 224439      224538: gap of unknown length
* 224539      225707: contig of 1169 bp in length
* 225708      225807: gap of unknown length
* 225808      227705: contig of 1898 bp in length.
```

FEATURES

source

```
1..227705
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9613"
/clone="CH240-15N19"
3270..3821
/estimated_length=552
6292..6500
/estimated_length=209
10079..10128
/estimated_length=50
20019..20068
/estimated_length=50
24149..24198
/estimated_length=50
26017..26656
/estimated_length=640
32725..32774
/estimated_length=50
46964..47013
/estimated_length=50
52516..52565
/estimated_length=50
62836..63012
/estimated_length=177
77339..77388
/estimated_length=50
81138..81187
/estimated_length=50
85198..85247
```

Query Match 83.2%; Score 20.8; DB 14; Length 227705;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY      2 AAAAAAAAAAGCATGACTGTGACAC 25
Db      173938 AAAAAAAAAATCATGACTGTGCCAC 173915
```

RESULT 45

AC120722

LOCUS AC120722 230172 bp DNA linear HTG 19-SEP-2002
DEFINITION Rattus norvegicus clone CH230-264X20, *** SEQUENCING IN PROGRESS
*** 3 unordered pieces.

AC120722 GI:23196173
VERSION AC120722.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 230172)
Muzny,D,Marle,,Metzker,M,Lea,,Abramson,S,,Adams,C,,Alder,J,,
Allen,C,,Allen,H,,Altschuler,S,,Amin,A,,Anguiano,D,,
Angelichsch,V,,Ayagi,A,,Ayodeji,M,,Baca,B,,Baden,H,,
Baldwin,D,,Bandarainake,D,,Barber,M,,Barnstead,M,,Benahmed,F,,

Blawie, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Butch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., d'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Ande, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Garcia, R., Garcia, A., Garner, T., Garza, M., Gebregregis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howell, S., Hulik, S., Hume, J., Idlebird, D., Jackson, L., Jackson, L., Jiang, H., Johnson, B., Johnson, R., Joilvet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C., Kwie, C., Kraf, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensueta, L., Louisedge, H., Lozado, R.J., Lu, X., Ma, J., Mathewswami, M., Mahindratne, M., Mahmood, M., Malloy, K., Mangum, B., Manjun, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Mlisoajevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundaasa, M., Murphy, M., Nakkevits, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nsoekemele, O., Okunolu, G., Olarnunagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pizzo, M., Quiroz, J., Rachlin, E., Reeves, K., Reiter, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, R., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, A., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Unpublished
Direct Submission
2 (bases 1 to 230172)

Worley, K.C.
Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 230172)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Sep 19, 2002 this sequence version replaced gi:21908529.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information

```

Center project name: GYVA
Center clone name: CH230-264K20
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 182947 bases at least Q40
Consensus quality: 185925 bases at least Q30
Consensus quality: 187995 bases at least Q20
Estimated insert size: 211479; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 83798: contig of 83798 bp in length
* 83799 83898: gap of unknown length
* 83899 228791: contig of 144893 bp in length
* 228792 228891: gap of unknown length
* 228892 230172: contig of 1281 bp in length.
*
Location/Qualifiers
1. 230172
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-264K20"
1. 1047
/note="wgs contig"
1098. 3155
/note="wgs contig"
83799. 83898
/estimated_length=unknown
83899. 85134
/note="wgs contig"
88902. 91277
/note="wgs contig"
228792. 228891
/estimated_length=unknown

ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 230172;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0.

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
|||||
|||||
AAAAAAAAAAGCATTTACTGTGACA 33080

Db 33057 AAAAAAAAAAGCATTTACTGTGACA 33080

RESULT 46
AC131436/c
LOCUS
DEFINITION
AC131436 236814 bp DNA linear HTG 09-MAY-2003
Rattus norvegicus clone CH230-5Pl, *** SEQUENCING IN PROGRESS ***,
3 unordered pieces.
AC131436
AC131436 GI:30467795
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 236814)
Muzny,D.,Marle, Metzker,M.,Lee, Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalto,K., Blair,J., Blankenhorn,K., Blyth,P., Brown,M.,

```

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Deedrich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Ebecato, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howell, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jang, J., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowik, C., Kraft, C.L., Lebrow, H., Levant, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzuela, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokwelu, O., Okunolu, G., Olarnpungsoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richard, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valae, R., Vera, V., Villanar, D., Waldron, J., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Niederhuser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 236814)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (22-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 236814)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced g1:24818187.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center
Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCHQ
Center clone name: CH230-5P1
----- Summary Statistics
Assembly program: Atlas
Consensus quality: 222197 bases at least Q40
Consensus quality: 224952 bases at least Q30
Consensus quality: 226289 bases at least Q20
Estimated insert size: 238764; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 233065: contig of 233065 bp in length
* 233066 233165: gap of unknown length
* 233166 234362: contig of 1197 bp in length
* 234363 234462: gap of unknown length
* 234463 236814: contig of 2352 bp in length.
Location/Qualifiers
1.236814
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-5P1"
199902..200779
/note="clone boundary
clone_end:Sp6
site:BCORI
end_sequence:BH363605"
complement(230476..231247)
/note="clone boundary
clone_end:T7
site:BCORI
end_sequence:BH363603"
231330..233065
/note="wgs end_extension
clone_end:T7"
233066..233165
/estimated_length=unknown
234363..234462
/estimated_length=unknown
ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 236814;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 101216 AAAAAAAAAAATGACTCTTACA 101193
1 AAAAAAAAAAAGCATGACTGTGACA 24
|||||
AAAAAAAAAAAAAAAAAATGACTCTTACA 101193
RESULT 47
AC162048 251434 bp DNA 1linear HTG 01-JUL-2005
LOCUS AC162048 Bos taurus clone CH240-111C3, *** SEQUENCING IN PROGRESS ***, 25
DEFINITION unbordered pieces.
ACCESSION AC162048
VERSION AC162048.2 GI:68302765
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

1 (bases 1 to 251434)

COMMENT

Munty, D., Martie, Metzger, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrechts, S., Amin, A., Arguiano, D., Araya, V., Ayoubi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bryant, C., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davis, M., Davis, C., Davy-Carroil, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, D., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzshwa, L., Louissege, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Mlosoavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munsadas, M., Murphy, M., Naik, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nsoekleheneh, O., Okwunonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaiz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steele, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmami, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL

Unpublished
2 (bases 1 to 251434)

REFERENCE
AUTHORS

Worley, K.C.

Submitted (25-MAY-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 251434)

REFERENCE
AUTHORS

Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jun 29, 2005 this sequence version replaced gi:66571366. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: FGXP

Center clone name: CH240-11C13

Summary Statistics

Assembly program: Atlas 3.0:

Consensus quality: 240429 bases at least Q40

Consensus quality: 244119 bases at least Q30

Consensus quality: 246537 bases at least Q20

Estimated insert size: 251354; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbankdraft_data.html)

NOTE: This sequence may represent more than one clone.

NOTE: This is a 'working draft' sequence. It currently

consists of 25 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1	44178	contig of 44178 bp in length
44179	44228	gap of 50 bp
44229	47927	contig of 3699 bp in length
47928	47977	gap of 50 bp
47978	67124	contig of 19147 bp in length
67125	67174	gap of 50 bp
67175	80890	contig of 13716 bp in length
80891	80940	gap of 50 bp
80941	90027	contig of 9087 bp in length
90028	90803	gap of 776 bp
90804	92372	contig of 1569 bp in length
92373	92472	gap of unknown length
92473	120485	contig of 28013 bp in length
120486	120535	gap of 50 bp
120536	145165	contig of 24630 bp in length
145166	145215	gap of 50 bp
145216	157706	contig of 12491 bp in length
157707	157756	gap of 50 bp
157757	160626	contig of 2870 bp in length
160627	160676	gap of 50 bp
160677	165242	contig of 4566 bp in length
165243	165292	gap of 50 bp
165293	171029	contig of 5737 bp in length
171030	171079	gap of 50 bp
171080	172978	contig of 1899 bp in length
172979	173028	gap of 50 bp
173029	176556	contig of 3528 bp in length
176557	176606	gap of 50 bp
176607	185103	contig of 8497 bp in length
185104	185153	gap of 50 bp
185154	192182	contig of 7029 bp in length
192183	192232	gap of 50 bp
192233	200381	contig of 8149 bp in length
200382	200431	gap of 50 bp
200432	202973	contig of 2542 bp in length
202974	203023	gap of 50 bp
203024	219071	contig of 16048 bp in length
219072	219121	gap of 50 bp
219122	238577	contig of 19456 bp in length

```

* 238578 238677: gap of unknown length
* 238678 240316: contig of 1639 bp in length
* 240317 240416: gap of unknown length
* 240417 241637: contig of 1221 bp in length
* 241638 241737: gap of unknown length
* 241738 243666: contig of 1929 bp in length
* 243667 243766: gap of unknown length
* 243767 246733: contig of 2967 bp in length
* 246734 251434: contig of 4601 bp in length.
* 246834

```

FEATURES

```

source
1. 251434
   /organism="Bos taurus"
   /mol_type="genomic DNA"
   /db_xref="taxon:9913"
   /clone="CH240-111C13"
   44179..44228
   /estimated_length=50
   47928..47977
   /estimated_length=50
   67125..67174
   /estimated_length=50
   80891..80940
   /estimated_length=50
   90028..90803
   /estimated_length=776
   92373..92472
   /estimated_length=unknown
   120486..120535
   /estimated_length=50
   145166..145215
   /estimated_length=50
   157707..157756
   /estimated_length=50
   160627..160676
   /estimated_length=50
   165243..165292
   /estimated_length=50
   171030..171079
   /estimated_length=50
   172979..173028
   /estimated_length=50
   176557..176606
   /estimated_length=50
   185104..185153
   /estimated_length=50

```

```

Query Match      83.2%  Score 20.8; DB 14; Length 251434;
Best Local Similarity 91.7%  Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 AAAAAAAAAAGCATGCTGTGCAC 25
Db 74414 AAAAAAAAAATCATGCTGTGCAC 74391

```

```

RESULT 48
AC123311/c 264197 bp DNA linear HTG 19-NOV-2002
LOCUS AC123311/c
DEFINITION Rattus norvegicus clone CH230-89N18, *** SEQUENCING IN PROGRESS
*** 2 unordered pieces.
AC123311 264197 bp DNA linear HTG 19-NOV-2002
AC123311 GI:25073543
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 264197)
Muzny,D,Marie,,Metzker,M,Lee,,Abramson,S,,Adams,C,,Alder,J,,
Allen,C,,Allen,H,,Aisbroot,S,,Amin,A,,Anguiano,D,,
Anyalebechi,V,,Aoyagi,A,,Ayodeji,M,,Baca,R,,Baden,H,,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
COMMENT

```

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bissal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Butcher, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Becotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabriel, A., Garcia, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, B., Geer, K., Gill, R., Grady, M., Guertler, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewe, L., Louieged, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackemebe, O., Okunolu, G., Olarunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindecker, A., Popovic, D., Primus, B., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shetman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorreller, R., Soosa, V., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K., Vales, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wlarczyk, R., Wooden, H., Woley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 264197)
Worley, K. C.

Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 264197)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23097381.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUGS
Center clone name: CH230-89N18
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 230334 bases at least Q40
Consensus quality: 233663 bases at least Q30
Consensus quality: 235400 bases at least Q20
Estimated insert size: 226450; sum-of-coverage estimation
Quality coverage: 6x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 10223: contig of 10223 bp in length
* 10224 10323: gap of unknown length
* 10324 264197: contig of 253874 bp in length.
-----
FEATURES
Source:
1. 264197 Location/Qualifiers
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-89N18"
    1. 1563
        /note="wgs_end_extension
        clone_end:T7"
        10224. 10323
        /estimated_length=unknown
        10324. 11502
        /note="wgs_end_extension
        clone_end:T7"
        complement (10470..11262)
        /note="clone_boundary
        clone_end:T7
        site:
        end_sequence: BH342569"
        complement (14141..246198)
        /note="clone_boundary
        clone_end:Sp6
        site:
        end_sequence: BH342570"
        256722. 258581
        /note="wgs_end_extension
        clone_end:Sp6"
        261440..264197
        /note="wgs_end_extension
        clone_end:Sp6"
-----
ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 264197;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
-----
Db 1 AAAAAAAAAAGCATGACTGTGACA 24
248113 AAAAAAAAAAGAACTGTTACA 248090
-----
RESULT 49
AC156526 AC156526 270106 bp DNA linear HTG 01-JUL-2005

```

```

DEFINITION
Bos taurus clone CH240-5402, *** SEQUENCING IN PROGRESS ***, 28
unordered pieces.
AC156526
AC156526.3 GI:68266773
HTG: HTGS_PHASE1, HTGS_DRAFT, HTGS_ENRICHED.
Bos taurus (cow)
Bos taurus (cow)
Bos taurus
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 270106)
REFERENCE
AUTHORS
Muzny, D., Maric, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Ayalebech, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyne, M., Cree, A., D'Souza, L.,
Devila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Gills, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hui, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowik, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louie, H., Lozano, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,
Milosavljevic, A., Miner, G., Mijang, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwakolam, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, D., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sores, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, J., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Yu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhou, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 270106)
Worley, K.C.
TITLE
JOURNAL
Direct Submission
Submitted (30-JAN-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 270106)
REFERENCE
AUTHORS
Cow Genome Sequencing Consortium.
TITLE
JOURNAL
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

COMMENT

Baylor Plaza, Houston, TX 77030, USA
 On Jun 28, 2005 this sequence version replaced gi:58801743.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: FDB3
 Center clone name: CH240-5402
 ----- Summary Statistics

Assembly program: Atlas 3.0
 Consensus quality: 247510 bases at least Q40
 Consensus quality: 250598 bases at least Q30
 Consensus quality: 253240 bases at least Q20
 Estimated insert size: 254020; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 12362: contig of 12362 bp in length
 12363 12720: gap of 358 bp
 12721 18283: contig of 5563 bp in length
 18284 18333: gap of 50 bp
 18334 53856: contig of 35523 bp in length
 53857 53906: gap of 50 bp
 53907 70984: contig of 17078 bp in length
 70985 71084: gap of unknown length
 71085 97779: contig of 26695 bp in length
 97780 97829: gap of 50 bp
 97830 99537: contig of 1708 bp in length
 99538 99587: gap of 50 bp
 99588 110398: contig of 10811 bp in length
 110399 110448: gap of 50 bp
 110449 125359: contig of 14911 bp in length
 125360 125409: gap of 50 bp
 125410 139015: contig of 13606 bp in length
 139016 139065: gap of 50 bp
 139066 140532: contig of 1467 bp in length
 140533 140582: gap of 50 bp
 140583 142781: contig of 2199 bp in length
 142782 142851: gap of 70 bp
 142852 145589: contig of 2738 bp in length
 145590 145639: gap of 50 bp
 145640 155352: contig of 9713 bp in length
 155353 155402: gap of 50 bp
 155403 170548: contig of 15146 bp in length
 170549 170598: gap of 50 bp
 170599 179538: contig of 8940 bp in length
 179539 179638: gap of unknown length
 179639 187668: contig of 8030 bp in length
 187669 187718: gap of 50 bp

FEATURES

* 187719 194430: contig of 6712 bp in length
 * 194431 195737: gap of 1307 bp
 * 195738 217195: contig of 21458 bp in length
 * 217196 217245: gap of 50 bp
 * 217246 220054: contig of 2809 bp in length
 * 220055 220154: gap of unknown length
 * 220155 233149: contig of 12995 bp in length
 * 233150 234767: gap of 50 bp
 * 234768 245110: contig of 1568 bp in length
 * 245111 248243: contig of 3133 bp in length
 * 248244 248343: gap of unknown length
 * 248344 249357: contig of 1014 bp in length
 * 249358 249457: gap of unknown length
 * 249458 250903: contig of 1446 bp in length
 * 250904 251003: gap of unknown length
 * 251004 253321: contig of 2318 bp in length
 * 253322 253421: gap of unknown length
 * 253422 254795: contig of 1374 bp in length
 * 254796 254895: gap of unknown length
 * 254896 262736: contig of 7841 bp in length
 * 262737 262836: gap of unknown length
 * 262837 270106: contig of 7270 bp in length.

source

1..270106
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"
 /clone="CH240-5402"

gap 12363..12720
 /estimated_length=358
 gap 18284..18333
 /estimated_length=50
 gap 53857..53906
 /estimated_length=50
 gap 70985..71084
 /estimated_length=unknown
 gap 97780..97829
 /estimated_length=50
 gap 99538..99587
 /estimated_length=50
 gap 110399..110448
 /estimated_length=50
 gap 125360..125409
 /estimated_length=50
 gap 139016..139065
 /estimated_length=50
 gap 140533..140582
 /estimated_length=50
 gap 142782..142851
 /estimated_length=70
 gap 145590..145639
 /estimated_length=50

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 14; Length 270106;
 Matches 22; Conservativity 91.7%; Pred. No. 1.8e+02; Indels 0; Gaps 0;

1 AAAAAAAAAAGCATGACGTGACCA 24

Db 208470 AAAAAAAAAAGCATGACGTGACCA 208493

RESULT 50

AC129648

LOCUS

AC129648 280215 bp DNA linear HTG 13-NOV-2002
 Rattus norvegicus clone CH250-8H18, WORKING DRAFT SEQUENCE, 3
 unorderd pieces.

AC129648

AC129648.4 GI:24940664

VERSION

HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM Rattus norvegicus

REFERENCE
AUTHORS

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurungathi; Murioidea; Muridae; Murinae; Rattus.

1 (bases 1 to 280215)

Mizny,D,Marie, Metzger,M,lee, Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D., Anyalabechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalto,K., Blair,U., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Day-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Diya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gnatetue,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howell,S., Huliyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshuhwa,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmud,M., Mallyo,K., Mangum,A., Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,B., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Mlosoavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Naokoeleneh,O., Okunou,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfennkch,D., Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L., L., Piazzi,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richard,S., Riggs,F., Rives,C., Rodkey,T., Rojars,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soes,J., Steinhle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstein,G., and Gibbs,R.A.

TITLE
JOURNALREFERENCE
AUTHORS

Unpublished
2 (bases 1 to 280215)

TITLE
JOURNAL

Submitted (31-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE
JOURNAL

COMMENT

On Nov 13, 2002 this sequence version replaced gi:23096222. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDNWJ

Center clone name: CH230-8H18

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 237591 bases at least Q40

Consensus quality: 239484 bases at least Q30

Consensus quality: 240656 bases at least Q20

Estimated insert size: 24055; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 255140: contig of 255140 bp in length
* 255141 255240: gap of unknown length
* 255241 274427: contig of 19187 bp in length
* 274428 274527: gap of unknown length
* 274528 280215: contig of 5688 bp in length.

FEATURES

source

1. 280215

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-8H18"

1. 1128

/note="wgs contig"

misc_feature

1179..5584

/note="wgs contig"

255141..255240

/estimated_length=unknown

274428..274527

/estimated_length=unknown

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 280215;

Best Local Similarity 91.7%; Pred. No. 1.8e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAAAGCATGACTGTGACA 24

18483 AAAAAAAAAAGCATGACTGTGACA 18506

Search completed: December 14, 2005, 11:11:02
Job time : 879.8 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 ; Search time 203.2 Seconds
(without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-5
Perfect score: 25
Sequence: 1 aaaaaaaaaagcatgactgtgacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

N_Geneseq_21:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004as:*
- 13: geneseqn2004bs:*
- 14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	22.4	89.6	2013	3	AAC42808	Aac42808 Arabidops
C 2	22.4	89.6	2013	6	AB213714	Ab213714 Arabidops
C 3	20.2	80.6	4860	4	AAK65920	Aak65920 Human imm
C 4	19.8	79.2	1363	10	AD61476	Ad61476 Rat gene
C 5	19.8	79.2	1364	10	AD667617	Ad667617 Human Lyr1
C 6	19.8	79.2	217409	11	ACN45150	Actn45150 Human gen
C 7	19.4	77.6	32404	9	ADA02894	Ada02894 Human BLR
C 8	19.4	77.6	32404	10	ADB72632	Adb72632 Human BLR
C 9	19.4	77.6	32404	10	ADC85373	Adc85373 Mouse BLR
C 10	19.4	77.6	32404	12	ADM74489	Adm74489 Human car
C 11	19.4	77.6	32404	13	ADR66970	Adr66970 Human can
C 12	19.4	77.6	33296	14	AD212715	Ad212715 Human can
C 13	19.2	76.8	1139	2	AAT79693	Aat79693 BRCA2 can
C 14	19.2	76.8	2392	10	ADR82456	Adr82456 Leukemia
C 15	19.2	76.8	2679	4	AA503718	Aa503718 DNA encod
C 16	19.2	76.8	3956	12	ADQ22842	Adq22842 Human sof
C 17	19.2	76.8	5830	2	AA237153	Aa237153 DNA segue
C 18	19.2	76.8	7764	5	AD163336	Ad163336 Human ova
C 19	19.2	76.8	7764	10	AAD54628	Aad54628 Human chr

C 20	19.2	76.8	7764	12	ADK60458	Adk60458 Angiogene
C 21	19.2	76.8	7764	12	ADK60759	Adk60759 Angiogene
C 22	19.2	76.8	7764	12	ADP73081	Adp73081 Angiogene
C 23	19.2	76.8	7764	14	ADX44407	Adx44407 Human chr
C 24	19.2	76.8	7764	14	ADY39318	Ady39318 Human CHD
C 25	19.2	76.8	17757	4	AAK80953	Aak80953 Human imm
C 26	19.2	76.8	28564	10	ADD48759	Add48759 Human gen
C 27	18.8	75.2	579	3	AAC35631	Aac35631 Arabidops
C 28	18.8	75.2	1659	13	AD54464	Ad54464 Plant ful
C 29	18.8	75.2	2613	4	ABL17270	Ab117270 Drosophil
C 30	18.8	75.2	44063	13	ABD33532	Abd33532 Human can
C 31	18.8	75.2	81440	6	ABQ76617	Abq76617 C. albica
C 32	18.8	75.2	110000	5	AA161373_2	Continuation (3 of
C 33	18.6	74.4	361	5	ABV01013	Abv01013 Human pro
C 34	18.6	74.4	396	5	ABV10182	Abv10182 Human pro
C 35	18.6	74.4	415	5	ABV13353	Abv13353 Human pro
C 36	18.6	74.4	415	5	ABV40322	Abv40322 Human pro
C 37	18.6	74.4	415	5	ABV43675	Abv43675 Human pro
C 38	18.6	74.4	561	13	ACN59204	Actn59204 Cotton gy
C 39	18.6	74.4	686	5	ABV26507	Abv26507 Human pro
C 40	18.6	74.4	686	5	ABV20665	Abv20665 Human pro
C 41	18.6	74.4	686	5	ABV22901	Abv22901 Human pro
C 42	18.6	74.4	686	5	ABV28735	Abv28735 Human pro
C 43	18.6	74.4	711	12	ADQ25229	Adq25229 Human sof
C 44	18.6	74.4	1194	6	ABN69451	Abn69451 Streptoco
C 45	18.6	74.4	1268	13	ADX13182	Adx13182 Plant ful
C 46	18.6	74.4	1612	10	AD61980	Ad61980 Rat gene
C 47	18.6	74.4	1612	10	AD661984	Ad661984 Rat gene
C 48	18.6	74.4	3216	13	ADY97351	Ady97351 N tabacum
C 49	18.6	74.4	5020	4	AAH18707	Aah18707 Human CDN
C 50	18.6	74.4	6495	4	AA106441	Aa106441 Human rep
C 51	18.6	74.4	6495	5	AA540553	Aa540553 DNA encod
C 52	18.6	74.4	6495	11	ADJ09759	Adj09759 Human pro
C 53	18.6	74.4	7856	14	ADX58003	Adx58003 Transfer
C 54	18.6	74.4	8127	14	ADX58004	Adx58004 Transfer
C 55	18.6	74.4	9189	14	ADK57926	Adk57926 Caprine a
C 56	18.6	74.4	14103	4	AAK61278	Aak61278 Human exc
C 57	18.6	74.4	14103	5	AA163700	Aa163700 Human kid
C 58	18.6	74.4	14103	5	ABL27058	Ab127058 Drosophil
C 59	18.6	74.4	14779	9	ADN02495	Adn02495 Mouse Myc
C 60	18.6	74.4	25032	10	ADB72233	Adb72233 Mouse Myc
C 61	18.6	74.4	25032	10	ADB82935	Adb82935 Mouse Myc
C 62	18.6	74.4	25032	10	ADB82935	Adb82935 Mouse Myc
C 63	18.6	74.4	25032	10	ADB82935	Adb82935 Mouse Myc
C 64	18.6	74.4	35057	8	AA151501	Aa151501 Human pan
C 65	18.6	74.4	59065	6	AB142416	Ab142416 Human ser
C 66	18.6	74.4	59065	6	ABD1839	Abd1839 Human LIM
C 67	18.6	74.4	59065	10	ADG98727	Adg98727 Human LIM
C 68	18.6	74.4	59065	13	ADR44808	Adr44808 Human kin
C 69	18.6	74.4	60815	11	ACN43882	Actn43882 Human gen
C 70	18.6	74.4	93323	12	ADQ97047_3	Adq97047_3
C 71	18.6	74.4	113079	13	ABD32998	Abd32998 Human can
C 72	18.6	74.4	128978	6	ABK83459	Abk83459 Human CDN
C 73	18.6	74.4	128978	8	AA054587	Aa054587 Human LIM
C 74	18.6	74.4	128978	13	ADR52994	Adr52994 Drug cher
C 75	18.6	74.4	128978	14	ADX07226	Adx07226 Cyclin-de
C 76	18.6	74.4	160552	4	AA002697	Aa002697 Human gly
C 77	18.6	74.4	163321	11	ACN43898	Actn43898 Human gen
C 78	18.6	74.4	164841	11	ACN44428	Actn44428 Mouse gen
C 79	18.6	74.4	276276	11	ACN44350	Actn44350 Human gen
C 80	18.4	73.6	1298	13	ADX30404	Adx30404 Plant ful
C 81	18.4	73.6	4397	12	ADQ95667	Adq95667 T cell ac
C 82	18.4	73.6	44567	9	AA056117	Aa056117 Human BAT
C 83	18.4	73.6	44567	9	ADA02479	Ada02479 Human BAT
C 84	18.4	73.6	44567	10	ADB72218	Adb72218 Human BAT
C 85	18.2	72.8	41	10	ADC36978	Adc36978 F11-in o
C 86	18.2	72.8	41	14	AD200545	Ad200545 Fragment
C 87	18.2	72.8	362	4	ABA56464	Ab56464 Human foe
C 88	18.2	72.8	362	4	ABA56464	Ab56464 Human foe
C 89	18.2	72.8	362	4	ABA56464	Ab56464 Human foe
C 90	18.2	72.8	362	4	ABA45937	Ab45937 Human bre
C 91	18.2	72.8	362	4	ABA26097	Ab26097 Probe #45
C 92	18.2	72.8	362	4	AAK30134	Aak30134 Human bon

C 93	18.2	72.8	362	4	AAK04673
C 94	18.2	72.8	362	4	ABE29784
C 95	18.2	72.8	362	5	AA104536
C 96	18.2	72.8	362	6	ABSO4715
C 97	18.2	72.8	443	13	ACP82377
C 98	18.2	72.8	486	9	ACH23577
C 99	18.2	72.8	560	12	ACH76292
C 100	18.2	72.8	771	4	AAS30431
C 101	18.2	72.8	771	4	AAS30422
C 102	18.2	72.8	771	4	AAS30420
C 103	18.2	72.8	771	4	AA104453
C 104	18.2	72.8	771	4	AA104454
C 105	18.2	72.8	771	4	AA104455
C 106	18.2	72.8	792	4	AAK83879
C 107	18.2	72.8	792	4	AAK83878
C 108	18.2	72.8	1267	12	AD035984
C 109	18.2	72.8	2000	11	AC136496
C 110	18.2	72.8	2402	6	ABE76551
C 111	18.2	72.8	2487	6	ABE76558
C 112	18.2	72.8	2543	6	ABK34570
C 113	18.2	72.8	2915	12	AD057552
C 114	18.2	72.8	3388	4	AAE64637
C 115	18.2	72.8	4572	4	AAE64636
C 116	18.2	72.8	4597	4	AAE64632
C 117	18.2	72.8	4628	4	AAE64631
C 118	18.2	72.8	6008	4	AB119584
C 119	18.2	72.8	12118	6	ABK90083
C 120	18.2	72.8	12118	13	ACR40377
C 121	18.2	72.8	48459	11	ACN44130
C 122	18.2	72.8	49914	13	ABD23362
C 123	18.2	72.8	52754	9	ADAO2788
C 124	18.2	72.8	52754	10	ABE72536
C 125	18.2	72.8	52754	10	AD085278
C 126	18.2	72.8	52754	12	ADM74393
C 127	18.2	72.8	62444	12	AD097574
C 128	18.2	72.8	65274	14	AD212653
C 129	18.2	72.8	68582	13	ABD33059
C 130	18.2	72.8	110000	2	AA123053_14
C 131	18.2	72.8	110000	2	AAV21209_00
C 132	18.2	72.8	110000	11	ACN43984_3
C 133	18.2	72.8	110000	11	ACN43984_3
C 134	18.2	72.8	128034	10	ADBE43582
C 135	18.2	72.8	128034	10	ADBE43581
C 136	18.2	72.8	128034	12	ADBE4359
C 137	18.2	72.8	128034	8	ADBE4357
C 138	18.2	72.8	177380	8	ADBE20870
C 139	18.2	72.8	177380	8	ADBE20870
C 140	18.2	72.8	177380	10	ADBE96442
C 141	18.2	72.8	177380	10	ADBE92133
C 142	18.2	72.8	177380	10	ADBE92133
C 143	18.2	72.8	202100	10	ADBE43515
C 144	18.2	72.8	202100	12	ADBE4357
C 145	18	72.0	137908	11	ADBE56534
C 146	17.8	71.2	329	6	AB181814
C 147	17.8	71.2	356	6	AB164629
C 148	17.8	71.2	356	6	AB180674
C 149	17.8	71.2	377	6	AB187666
C 150	17.8	71.2	381	5	ABV10042
C 151	17.8	71.2	381	5	ABV10042

ALIGNMENTS

RESULT	1
ID	AAC42808/c
XX	AAC42808 standard; DNA; 2013 BP.
XX	
AC	
XX	AAC42808;
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 36926
XX	

KW	Hybridization assay; genetic mapping; gene expression control; protein identification; signal transduction pathway; metabolic pathway
KM	promoter; termination sequence; ss.
XX	
OS	Arabidopsis thaliana.
XX	
XX	
FN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000BP-00301439.
XX	
PR	25-FEB-1999; 99US-0121825P.
PR	05-MAR-1999; 99US-0123180P.
PR	09-MAR-1999; 99US-0123548P.
PR	23-MAR-1999; 99US-0125788P.
PR	25-MAR-1999; 99US-0126264P.
PR	29-MAR-1999; 99US-0126785P.
PR	01-APR-1999; 99US-0127462P.
PR	06-APR-1999; 99US-0128234P.
PR	16-APR-1999; 99US-0128714P.
PR	18-APR-1999; 99US-0129845P.
PR	19-APR-1999; 99US-0130077P.
PR	21-APR-1999; 99US-0130442P.
PR	23-APR-1999; 99US-0130510P.
PR	23-APR-1999; 99US-0130891P.
PR	28-APR-1999; 99US-0131449P.
PR	30-APR-1999; 99US-0133048P.
PR	30-APR-1999; 99US-0133207P.
PR	04-MAY-1999; 99US-0132484P.
PR	05-MAY-1999; 99US-0132485P.
PR	06-MAY-1999; 99US-0132486P.
PR	07-MAY-1999; 99US-0132487P.
PR	11-MAY-1999; 99US-0132863P.
PR	14-MAY-1999; 99US-0134256P.
PR	14-MAY-1999; 99US-0134216P.
PR	14-MAY-1999; 99US-0134219P.
PR	14-MAY-1999; 99US-0134221P.
PR	14-MAY-1999; 99US-0134370P.
PR	18-MAY-1999; 99US-0134768P.
PR	19-MAY-1999; 99US-0134941P.
PR	20-MAY-1999; 99US-0135124P.
PR	21-MAY-1999; 99US-0135353P.
PR	24-MAY-1999; 99US-0135629P.
PR	25-MAY-1999; 99US-0136021P.
PR	27-MAY-1999; 99US-0136392P.
PR	28-MAY-1999; 99US-0136782P.
PR	01-JUN-1999; 99US-0137222P.
PR	03-JUN-1999; 99US-0137528P.
PR	04-JUN-1999; 99US-0137502P.
PR	07-JUN-1999; 99US-0137724P.
PR	08-JUN-1999; 99US-0138094P.
PR	10-JUN-1999; 99US-0138540P.
PR	10-JUN-1999; 99US-0138847P.
PR	14-JUN-1999; 99US-0139119P.
PR	16-JUN-1999; 99US-0139452P.
PR	16-JUN-1999; 99US-0139453P.
PR	17-JUN-1999; 99US-0139492P.
PR	18-JUN-1999; 99US-0139454P.
PR	18-JUN-1999; 99US-0139455P.
PR	18-JUN-1999; 99US-0139456P.
PR	18-JUN-1999; 99US-0139462P.
PR	18-JUN-1999; 99US-0139463P.
PR	18-JUN-1999; 99US-0139750P.
PR	18-JUN-1999; 99US-0139763P.
PR	21-JUN-1999; 99US-0139817P.
PR	22-JUN-1999; 99US-0139899P.
PR	23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140685P.
PR 26-JUN-1999; 99US-0140882P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142330P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142920P.
PR 13-JUL-1999; 99US-0142977P.
PR 14-JUL-1999; 99US-0143542P.
PR 15-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145122P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.

PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158332P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161040P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 89.6%; Score 22.4; DB 3; Length 2013;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCTGACGTGACA 24
DB 1329 AAAAAACAGACGTGACGTGACA 1306

RESULT 2
ABZ13714/c
ID ABZ13714 standard; DNA; 2013 BP.
XX ABZ13714;
XX 21-JAN-2003 (first entry)
XX Arabidopsis thaliana stress regulated gene SRQ ID NO 1519.
DE Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX WO200216655-A2.
PN XX
XX 28-FEB-2002.
PD

XX 24-AUG-2001; 2001WO-US026685.
PF 24-AUG-2000; 2000US-0227866P.
XX 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX (SCRT) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Harper JF, Kreps J, Wang X, Zhu T;
PI WPI; 2002-304127/34.
XX Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
PS Claim 144; SEQ ID NO 1519; 577pp + Sequence Listing; English.
XX The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response; The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 2013 BP; 545 A; 384 C; 490 G; 594 T; 0 U; 0 Other;
Query Match 89.6%; Score 22.4; DB 6; Length 2013;
Best Local Similarity 95.8%; Pred. No. 25;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGCATGACTGTGACA 24
1329 AAAAAACAAGCATGACTGTGACA 1306
Db
RESULT 3
AAK69520
ID AAK69520 standard; DNA; 4860 BP.
XX
AC AAK69520;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24332.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
OS
XX WO200157182-A2.
PN
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228927P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231245P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235844P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 12-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.

PR	20-OCT-2000	2000US-0241786P	PR
PR	20-OCT-2000	2000US-0241787P	PR
PR	20-OCT-2000	2000US-0241808P	PR
PR	20-OCT-2000	2000US-0241809P	PR
PR	20-OCT-2000	2000US-0241826P	PR
PR	01-NOV-2000	2000US-0244617P	PR
PR	08-NOV-2000	2000US-0246474P	PR
PR	08-NOV-2000	2000US-0246475P	PR
PR	08-NOV-2000	2000US-0246476P	PR
PR	08-NOV-2000	2000US-0246477P	PR
PR	08-NOV-2000	2000US-0246538P	PR
PR	08-NOV-2000	2000US-0246539P	PR
PR	08-NOV-2000	2000US-0246524P	PR
PR	08-NOV-2000	2000US-0246525P	PR
PR	08-NOV-2000	2000US-0246526P	PR
PR	08-NOV-2000	2000US-0246527P	PR
PR	08-NOV-2000	2000US-0246528P	PR
PR	08-NOV-2000	2000US-0246532P	PR
PR	08-NOV-2000	2000US-0246609P	PR
PR	08-NOV-2000	2000US-0246610P	PR
PR	08-NOV-2000	2000US-0246611P	PR
PR	08-NOV-2000	2000US-0246613P	PR
PR	17-NOV-2000	2000US-0249207P	PR
PR	17-NOV-2000	2000US-0249208P	PR
PR	17-NOV-2000	2000US-0249209P	PR
PR	17-NOV-2000	2000US-0249210P	PR
PR	17-NOV-2000	2000US-0249211P	PR
PR	17-NOV-2000	2000US-0249212P	PR
PR	17-NOV-2000	2000US-0249213P	PR
PR	17-NOV-2000	2000US-0249214P	PR
PR	17-NOV-2000	2000US-0249215P	PR
PR	17-NOV-2000	2000US-0249216P	PR
PR	17-NOV-2000	2000US-0249217P	PR
PR	17-NOV-2000	2000US-0249218P	PR
PR	17-NOV-2000	2000US-0249244P	PR
PR	17-NOV-2000	2000US-0249245P	PR
PR	17-NOV-2000	2000US-0249264P	PR
PR	17-NOV-2000	2000US-0249265P	PR
PR	17-NOV-2000	2000US-0249297P	PR
PR	17-NOV-2000	2000US-0249299P	PR
PR	17-NOV-2000	2000US-0249300P	PR
PR	01-DEC-2000	2000US-0250160P	PR
PR	01-DEC-2000	2000US-0250391P	PR
PR	05-DEC-2000	2000US-0251030P	PR
PR	05-DEC-2000	2000US-0251988P	PR
PR	05-DEC-2000	2000US-0256719P	PR
PR	06-DEC-2000	2000US-0251479P	PR
PR	08-DEC-2000	2000US-0251566P	PR
PR	08-DEC-2000	2000US-0251688P	PR
PR	08-DEC-2000	2000US-0251869P	PR
PR	08-DEC-2000	2000US-0251899P	PR
PR	08-DEC-2000	2000US-0251990P	PR
PR	11-DEC-2000	2000US-0254097P	PR
PR	05-JAN-2001	2001US-0259678P	PR
XX	(HUMA-) HUMAN GENOME SCI INC.		XX
XX			XX
PI	Rosen CA, Barash SC, Ruben SM,		PI
DR	WPI, 2001-483426/52.		DR
XX			XX
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		PT
XX	useful for preventing, diagnosing and/or treating cancers and metastasis		XX
XX			XX
XX	Disclosure, SEQ ID NO 24332; 3071bp + Sequence Listing; English.		XX
AS	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)		AS
CC	amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic		CC
CC	activity, and can be used in gene therapy and vaccine production. (I)		CC
CC	proteins and polynucleotides may be used in the prevention, diagnosis and		CC
CC	treatment of diseases associated with inappropriate (I) expression. For		CC
CC	example, they may be used to treat disorders associated with decreased		CC
CC	expression by rectifying mutations or deletions in a patient's genome		CC

that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic-related diseases, especially cancers and cancer metastases of hematopoietic-derived cells. AAK6703 to AAK6769 represent human immune/hematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAM82169 represent sequences used in the exemplification of the present invention

Sequence 4860 BP; 1167 A; 1161 C; 1382 G; 1149 T; 0 U; 1 Other;

Query Match	80.8%;	Score 20.2;	DB 4;	Length 4860;
Best Local Similarity	88.0%;	Pred. No. 2.1e+02;		
Matches 22;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0

1 AAAAAAAAAAGCATGCTGTGCAC 25
|||||
Db 2101 AAAAAAAAAAGCATGCTGAAAC 2125

RESULT 4
ADE61476/c
ID ADE61476 standard; DNA; 403 BP.
AC ADE61476;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat gene AI639176, SEQ ID NO 7396.
XX
KM Raf; dg; gene; pain; neuronal tissue; gene therapy;
KM spinal segmental nerve injury; chronic constriction injury; CCI;
KM spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
PI Woolf C, D'Urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; AF639176.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
PS Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which is differentially expressed during pain. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPRO at
CC ftp.wipro.int/pub/published_pct_sequences.
XX
SQ Sequence 403 BP; 155 A; 63 C; 59 G; 126 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 10; Length 403;
Best Local Similarity 91.3%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 373 AAAAAAAAAAGCATGACTGTGAC 351

RESULT 5
ACN45150 standard; cDNA; 1364 BP.
ID ADD67617/c
XX
AC ADD67617;
XX
XX 15-JAN-2004 (first entry)
XX
DB Human LY1680P partial cDNA SEQ ID NO:94.
XX
XX haematological malignancy; immunocjugate; cytostatic; immunostimulant;
XX vaccine; immunotherapy; cancer; multiple myeloma cell;
XX chronic lymphocytic leukaemia; B cell leukaemia; lymphoma; anti-cancer;
XX human; gene; ss.
XX
OS Homo sapiens.
XX
XX WO2003062401-A2.
XX
XX 31-JUL-2003.
XX
XX 22-JAN-2003; 2003WO-US002353.
XX
XX 22-JAN-2002; 2002US-00057475.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Algate PA, Mannion J, Clapper JD, Wang A, Ordonez N;
XX Carter L, McNeill PD;
XX WPI; 2003-598749/56.
XX
XX New hematological malignancy-related genes and polypeptides, useful for
XX screening anti-cancer agents, and generating antibodies or
XX immunocjugates for treating e.g. multiple myeloma cell or chronic
XX lymphocytic leukemia.
XX
XX Claim 1; SEQ ID NO 94; 307bp; English.
XX
XX The present invention describes an isolated polynucleotide (1), which is
XX overexpressed in hematological malignancies, and which encodes a
XX polypeptide or an immunogenic fragment of the polypeptide. Also
XX described: (1) an isolated polypeptide; (2) an expression vector
XX comprising (1) operably linked to an expression control sequence; (3) a
XX host cell comprising an expression vector; (4) an isolated antibody that
XX specifically binds to the polypeptide or its immunogenic fragment; and

CC (5) immunocjugates comprising the antibody above, or an antibody that
CC specifically binds to a polypeptide, or its immunogenic fragment, encoded
CC by (1). (1) has cytostatic and immunostimulant activities, and can be
CC used in vaccines and immunotherapy. The immunocjugates are useful in
CC the manufacture of a medicament, particularly as active ingredients in a
CC composition for treating cancer, e.g. multiple myeloma cell, chronic
CC lymphocytic leukaemia, B cell leukaemia, or lymphomas in humans, sheep,
CC primates, goats, bovines, equines, porcines, lupines, canines or felines.
CC The polynucleotide (1) or polypeptide can be used for screening anti-
CC cancer agents, and generating antibodies or immunocjugates for treating
CC or preventing the above-mentioned diseases. The polynucleotide,
CC polypeptide or antibody can be used for detecting, diagnosing or
CC prognosticating the haematological malignancies described above. The
CC present sequence is used in the exemplification of the present invention.
XX
SQ Sequence 1364 BP; 396 A; 265 C; 300 G; 402 T; 0 U; 1 Other;

Query Match 79.2%; Score 19.8; DB 10; Length 1364;
Best Local Similarity 91.3%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAAAAAAAAGCATGACTGTGACAC 25
Db 437 AAAAAAAAAAGCATGACTGTGACAC 415

RESULT 6
ACN45150 standard; DNA; 217409 BP.
ID ACN45150
XX
XX ACN45150;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human genomic sequence hCG34092.
XX
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
XX OS Homo sapiens.
XX
XX WO2003073826-A2.
XX
XX 12-SEP-2003.
XX
XX 28-FEB-2003; 2003WO-US006235.
XX
XX 01-MAR-2002; 2002US-00087192.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
XX WPI; 2003-328604/31.
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
XX
XX Claim 1; SEQ ID NO 1954; 0bp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182566A1, for which no sequence data was published

XX Sequence 217409 BP; 59014 A; 45305 C; 46674 G; 64606 T; 0 U; 1810 Other;

Query Match 79.2%; Score 19.8; DB 11; Length 217409;
Best Local Similarity 91.3%; Pred. No. 4.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGAC 23
DB 5370 AAAAAAAAAAGCATGACTGGGC 5392

RESULT 7

ADA02894
ID ADA02894 standard; DNA; 32404 BP.

AC ADA02894;

DT 06-NOV-2003 (first entry)

DE Human BLR1 carcinoma associated gene, SEQ ID NO:1412.

XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;

KM prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;

OS Homo sapiens.

PN MO2003057146-A2.

PD 17-JUL-2003.

PR 26-DEC-2002; 2002WO-US041414.

PR 26-DEC-2001; 2001US-00035832.

XX (SAGR-) SAGRES DISCOVERY.

PI Morris DW;

DR WPI; 2003-587068/55.

PT New recombinant nucleic acid encoding carcinoma associated protein,
useful for preparing compositions for treating carcinomas.

PS Claim 1; SEQ ID NO 1412; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;

Query Match 77.6%; Score 19.4; DB 9; Length 32404;

Best Local Similarity 95.2%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
DB 302 AAAAAAAAAAGCATGATTGTG 322

RESULT 8

ADB72632
ID ADB72632 standard; DNA; 32404 BP.

AC ADB72632;

DT 04-DEC-2003 (first entry)

DE Human BLR1 gene.

XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;

KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

OS Homo sapiens.

PN MO2003008583-A2.

PD 30-JAN-2003.

PR 26-DEC-2001; 2001WO-US051291.

PR 02-MAR-2001; 2001US-00798586.

PR 23-OCT-2001; 2001US-00004113.

PR 08-NOV-2001; 2001US-00052482.

PR 30-NOV-2001; 2001US-00977722.

PR 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

PI Morris DW, Engelhard EK;

DR WPI; 2003-239337/23.

PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
cancers, neoplasm, adenocarcinoma, or sarcomas.

PS Claim 1; SEQ ID NO 460; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytosstatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;

Query Match 77.6%; Score 19.4; DB 10; Length 32404;
Best Local Similarity 95.2%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
DB 302 AAAAAAAAAAGCATGATTGTG 322

RESULT 9

ADC85373
ID ADC85373 standard; DNA; 32404 BP.

AC ADC85373;

DT 01-JAN-2004 (first entry)

DE Mouse B1r1 coding sequence.
 XX
 XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
 KW secreted; transmembrane; intracellular; ds.
 XX
 OS Mus sp.
 XX
 PN WO2003045230-A2.
 XX
 PD 05-JUN-2003.
 XX
 PF 02-DEC-2002; 2002WO-US038582.
 XX
 PR 30-NOV-2001; 2001US-00997722.
 XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW, Engelhard EK;
 XX
 DR WPI; 2003-513603/48.
 XX
 PT New recombinant nucleic acid comprising a nucleotide sequence of any of
 PT the carcinoma-associated (CA) genes, useful for screening for drug
 PT candidates for diagnosing or treating carcinomas.
 XX
 PS Claim 1; SEQ ID NO 159; 983bp; English.
 XX
 CC The invention relates to a recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the fully defined carcinoma-
 CC associated (CA) genes from the 50 tables given in the specification. The
 CC CA proteins are secreted, transmembrane or intracellular proteins. The
 CC recombinant nucleic acids are useful for screening for drug candidates
 CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
 CC ADC85514 represent CA genes of the invention.
 XX
 SQ Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;
 Query Match 77.6%; Score 19.4; DB 10; Length 32404;
 Best Local Similarity 95.2%; Pred. No. 5.3e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTG 21
 Db 302 AAAAAAAAAAGCATGACTGTG 322

RESULT 10
 ADM74489
 ID ADM74489 standard; DNA; 32404 BP.
 XX
 AC ADM74489;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Human carcinoma associated (CA) nucleic acid #79.
 XX
 KW Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
 KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
 KW cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN US2004072154-A1.
 XX
 PD 15-APR-2004.
 XX
 PF 30-NOV-2001; 2001US-00997722.
 XX
 PR 22-DEC-2000; 2000US-00747377.
 PR 02-MAR-2001; 2001US-00798586.
 XX
 PA (MORR/) MORRIS D W.
 PA (ENGEL/) ENGELHARD E K.

XX
 PI Morris DW, Engelhard EK;
 XX
 DR WPI; 2004-328562/30.
 XX
 PT New carcinoma associated gene or protein, useful for preparing a
 PT composition for diagnosing or treating carcinoma e.g., leukemia or
 PT lymphoma.
 XX
 PS Claim 1; SEQ ID NO 160; 29pp; English.
 XX
 CC The invention relates to new recombinant nucleic acids. The invention
 CC also relates to a host cell comprising a recombinant nucleic acid or
 CC expression vector, an expression vector comprising a recombinant nucleic
 CC acid, a recombinant protein, a method of screening for drug candidates, a
 CC method of screening for a bioactive agent capable of binding to a
 CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
 CC method of screening for a bioactive agent capable of modulating the
 CC activity of a CAP, a method of evaluating the effect of a candidate
 CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
 CC the activity of a CAP, a method of treating carcinomas, a method of
 CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
 CC propensity to carcinoma. A method of evaluating the effect of a candidate
 CC carcinoma drug comprises administering the drug to a patient, removing a
 CC cell sample from the patient and determining alterations in the
 CC expression or activation of a gene comprising the nucleotide sequence. A
 CC method of diagnosing carcinoma comprises determining the expression of
 CC one or more genes comprising the nucleic acid sequence in a first tissue
 CC type of a first individual and comparing the expression of the gene from
 CC a second normal tissue type from the first individual or a second
 CC unaffected individual, where a difference in the expression indicates
 CC that the first individual has carcinoma. A method of inhibiting the
 CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
 CC carcinomas comprises administering to a patient an inhibitor of CAP.
 CC Neutralising the effect of a CAP comprises contacting an agent specific
 CC for the CAP. The polypeptide specifically binds to the protein encoded by
 CC the nucleic acid. It comprises an antibody that specifically binds to the
 CC protein encoded by the nucleic acid. The nucleic acids are useful for
 CC preparing a composition for diagnosing or treating carcinoma e.g.,
 CC leukaemia or lymphoma. This sequence represents a human carcinoma
 CC associated (CA) nucleic acid of the invention. Note: The sequence data
 CC for this patent did not form part of the printed specification but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;
 Query Match 77.6%; Score 19.4; DB 12; Length 32404;
 Best Local Similarity 95.2%; Pred. No. 5.3e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTG 21
 Db 302 AAAAAAAAAAGCATGACTGTG 322

RESULT 11
 ADR66970
 ID ADR66970 standard; DNA; 32404 BP.
 XX
 AC ADR66970;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human cancer associated gene genomic sequence SEQ ID NO:16.
 XX
 KW cancer; cancer associated nucleic acid; cancer associated gene;
 KW cancer associated protein; CAP; cytostatic; vaccine; gene therapy;
 KW lymphoma; leukaemia; human; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2004074321-A2.

XX 02-SEP-2004.
PD 17-FEB-2004; 2004WO-US005000.
PF 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 23-SEP-2003; 2003US-00669920.
PR 15-DEC-2003; 2003US-00737318.
PA (SAGR-) SAGRES DISCOVERY INC.
PI Morris DW, Malandro MS;
XX WPI; 2004-652915/63.
DR P-PSDB; ADR66972.
XX
XX New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
XX Claim 16; SEQ ID NO 16; 166pp; English.
PS
XX
CC The present invention describes an isolated cancer associated (CA)
CC nucleic acid (1). Also described: (1) an expression vector comprising (1)
CC ; (2) a host cell comprising (1) or the expression vector; (3) a
CC microarray for detecting a CA nucleic acid; (4) an isolated cancer
CC associated protein (CAP) polypeptide, encoded within an open reading
CC frame of a CA sequence; (5) an isolated antibody, or its antigen binding
CC fragment, that binds to the above polypeptide; (6) a hybridoma that
CC produces the above monoclonal antibody; (7) a pharmaceutical composition
CC comprising the above antibody and a pharmaceutical excipient; (8) a kit
CC for detecting cancer cells, comprising the (monoclonal) antibody
CC described above; (9) methods for diagnosing cancer or for detecting the
CC presence or absence of cancer cells in an individual; (10) a method for
CC inhibiting growth of cancer cells in an individual; (11) a method for
CC delivering a therapeutic agent to cancer cells in an individual; (12) an
CC electronic library comprising the above polynucleotide or polypeptide, or
CC their fragments; (13) methods of screening for anticancer activity or for
CC a bioactive agent capable of modulating the activity of a CAP; (14)
CC methods for detecting cancer associated with expression of a polypeptide
CC in a test cell sample, or with the presence of an antibody in a test
CC serum sample; (15) a method for treating cancers; and (16) a method for
CC inhibiting the expression of CA gene in a cell. The CA sequences have
CC cytostatic activity, and can be used in vaccines, and in gene therapy.
CC The composition and methods are useful for detecting, diagnosing,
CC preventing and treating cancers, especially lymphoma and leukemia. They
CC may also be used in screening for agents that modulate cancer. The
CC present sequence represents a cancer associated gene genomic DNA
CC sequence, which is used in the exemplification of the present invention.
XX
SQ Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;
S0

Query Match 77.6%; Score 19.4; DB 13; Length 32404;
Best Local Similarity 95.2%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
Db 302 AAAAAAAAAAGCATGACTGTG 322

RESULT 12
AD212715
ID AD212715 standard; DNA; 33296 BP.
XX
XX AD212715;
AC
XX
XX 16-JUN-2005 (first entry)
XX
XX Human cancer-associated genomic DNA #19.
XX
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KM

KM cytostatic; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO2005031001-A2.
PN
XX
XX 07-APR-2005.
PD
XX
XX 23-SEP-2004; 2004WO-US031617.
PF
XX
XX 23-SEP-2003; 2003US-00669920.
PR
XX
XX (CHIR) CHIRON CORP.
PA
XX
XX Morris DW, Malandro MS;
PI
XX
XX WPI; 2005-273395/28.
DR
XX
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
PT comprises two or more nucleic acid probes.
PT
XX
XX
XX Disclosure; SEQ ID NO 235; 198pp; English.
PS
XX
CC The invention relates to a nucleic acid array for detecting a cancer
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC The invention also relates to a peptide array comprising two or more
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC that binds to a polypeptide, an isolated antibody or its fragment which
CC binds to a polypeptide, which is prepared by immunizing a host animal
CC with a composition comprising the polypeptide or its antigen binding
CC fragment and collecting cells from the host expressing antibodies against
CC the antigen or its antigen binding fragment, a composition comprising the
CC antibody and a carrier, a method of screening for anticancer activity, a
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC method of treating cancer and a method of inhibiting expression of a CA
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC nucleic acids. The antibody is useful for detecting the presence or
CC absence of cancer cells in an individual which involves contacting cells
CC from the individual with the antibody and detecting a complex of a CA
CC protein from the cancer cells and the antibody, where the detection of
CC the complex correlates with the presence of cancer cells in the
CC individual. The composition is useful for inhibiting growth of cancer
CC cells in an individual or for delivering a therapeutic agent to cancer
CC cells in an individual. The invention is also useful for diagnosing
CC cancer, for treating cancer and for inhibiting expression of a CA gene in
CC a cell. This sequence represents human cancer-associated genomic DNA of
CC the invention.
XX
SQ Sequence 33296 BP; 7974 A; 8525 C; 9288 G; 7509 T; 0 U; 0 Other;
S0

Query Match 77.6%; Score 19.4; DB 14; Length 33296;
Best Local Similarity 95.2%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
Db 676 AAAAAAAAAAGCATGACTGTG 696

RESULT 13
AAT79693/c
ID AAT79693 standard; DNA; 1139 BP.
XX
XX AAT79693;
AC
XX
XX 03-FEB-1998 (first entry)
XX
XX BRCA2 cancer susceptibility gene exon 24 and intron boundaries.
XX
XX BRCA2 cancer susceptibility gene; breast cancer; ovarian cancer;
KM gene therapy; prostate cancer; colorectal cancer; ocular melanoma;
KM leukemia; human; 88.
XX

		Solanum tuberosum. Synthetic.
OS	Key	Location/Qualifiers
XX	FH CDS	323..2680 /*tag= a
FT	/partial	
FT	/product= "Modified resistance protein 7"	
FT	/note= "The coding region only codes for amino acids 1 to 786 of the protein sequence shown in the specification and given in AAU02148"	
FT	/transl_except= (pos:806..808, aa:Glu)	
FT	/transl_except= (pos:850..852, aa:Thr)	
FT	/transl_except= (pos:853..855, aa:Thr)	
FT	/transl_except= (pos:974..976, aa:Thr)	
FT	/transl_except= (pos:983..985, aa:Glu)	
FT	/transl_except= (pos:1610..1612, aa:Thr)	
FT	/transl_except= (pos:2069..2070, aa:Val)	
FN	/note= "This codon contains an apparent 1 nucleotide deletion which alters the reading frame"	
PX	WO200129239-A2.	
PD	26-APR-2001.	
PF	12-OCT-2000; 2000WO-GB003930.	
PR	15-OCT-1999; 99GB-00024483.	
PA	(PLANT-) PLANT BIOSCIENCE LTD.	
P1	Bendahmane A., Baulcombe DC;	
D1	WPI; 2001-290924/30.	
DR	P-PADB; AAU02148.	
PT	Modifying activation characteristics of plant resistance proteins to produce autoactivator polypeptide capable of activation in absence of elicitor, by introducing modifications in amino acid sequence of protein.	
PS	Disclosure; Page 45-55; 77pp; English.	
CC	The sequence represents the coding sequence of Rx clone 7, a modified resistance autoactivator. Rx encodes an NBS-LRR polypeptide having a nucleotide binding site (NBS), leucine rich repeat (LRR), and which mediates cellular response leading to pathogen resistance and/or cell death or dysfunction in response to an elicitor. The Rx was modified by introducing modification to the amino acid sequence to produce an autoactivator polypeptide, capable of activation in absence of an elicitor. Decoupling of resistance response from its natural elicitor is useful for developing novel pathogen resistant plants. The modified CC resistance proteins are useful for conferring resistance to non-natural agents or stimuli and also for investigating resistance response pathways and protein interactions e.g. with activators and repressors	
QY	Sequence 2679 BP; 794 A; 475 C; 577 G; 833 T; 0 U; 0 Other; Query Match 76.8%; Score 19.2; DB 4; Length 2679; Best Local Similarity 87.5%; Pred. No. 4.9e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
D1	1 AAAAAGAAACATGCATGTGACA 24 125 AAAAGAAGAACGACACTCCTGACA 102	
RESULT 16		
ID	ADQ22842 standard; DNA; 3956 BP.	
AC	ADQ22842;	
XT	26-AUG-2004 (first entry)	

XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 5662.
DE	
XX	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX	ds.
KW	
KM	
XX	Homo sapiens.
OS	
XX	WO2004048938-A2.
PN	
PD	10-JUN-2004.
PP	
PF	26-NOV-2003; 2003WO-US038193.
XX	
PR	26-NOV-2002; 2002US-0429739P.
XX	(PROT-) PROTEIN DESIGN LABS INC.
PA	
PI	Aziz N, Ginsburg WM, Zlotnick A;
XX	WPI; 2004-441208/41.
DR	
XX	Early detection of soft tissue sarcoma comprises determining expression
PT	of a gene in a first soft tissue sample and a normal soft tissue sample
PT	and comparing the gene expression, also useful in treating soft tissue
PR	sarcoma.
XX	
PS	Example 2; SEQ ID NO 5662; 210pp; English.
XX	
CC	The invention relates to a novel method for detecting soft tissue sarcoma
CC	which comprises obtaining a first soft tissue sample from an individual
CC	and a normal soft tissue sample from the same or different individual,
CC	determining the expression of a gene in both samples and comparing the
CC	expression of the gene in both soft tissue samples, where a higher level
CC	of protein expression in the first soft tissue sample indicates the
CC	presence of soft tissue sarcoma. The method of the invention has
CC	cytostatic applications and may be useful for detecting soft tissue
CC	sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC	acid sequences may be useful in diagnostic and screening applications.
CC	The current sequence is that of a human soft tissue sarcoma-upregulated
CC	DNA of the invention. The current sequence is not shown within the
CC	specification per se but was submitted in CD format by the inventor.
XX	
SQ	Sequence 3956 BP; 1116 A; 821 C; 986 G; 1004 T; 0 U; 29 Other;
	Query Match 76.8%; Score 19.2; DB 12; Length 3956;
	Best Local Similarity 87.5%; Pred. No. 5.1e+02;
	Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
OY	1 AAAAAAAAAAGCATGACTGTGACA 24
DB	3298 AAAAAAAAAAGCATATCTGAGACA 3321
	RESULT 17
	AA237153/c
ID	AA237153 standard; DNA; 5820 BP.
XX	
AC	AA237153;
XX	
DT	01-FEB-2000 (first entry)
XX	
DE	DNA sequence of BAC77 including the potato Rx gene.
XX	
Potato; Rx gene; resistance gene; potato virus X; PVX; transgenic plant;	
broad spectrum extreme resistance; Narcissus mosaic virus; NMV; NVX; VMV;	
Nandina virus X; Viola mosaic virus; Cymbidium mosaic virus; CYMV; PopMV;	
Poplar mosaic virus; White clover mosaic virus; WClMV; activate; ss.	
XX	
SS	Solanum tuberosum.
XX	
Key	Location/Qualifiers
PH	2248..5407
CDS	.77

Db 6509 AAAAACAAGCAGACTTTGACAC 6486

RESULT 19
AAD54628/c
ID AAD54628 standard; DNA; 7764 BP.
XX
XX
AC AAD54628;
XX
XX 26-JUN-2003 (first entry)
DE Human chromodomain helicase DNA binding protein (CHD) encoding DNA #3.
XX
XX Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW cell proliferative disorder; chromatin organisation modifier domain;
KW cytosolic; ds.
XX
XX Homo sapiens.
XX
XX MO200298899-A2.
XX
XX 12-DEC-2002.
XX
XX 03-JUN-2002; 2002WO-US017466.
XX
XX 05-JUN-2001; 2001US-0296076P.
XX 10-OCT-2001; 2001US-0328605P.
XX 22-OCT-2001; 2001US-0338733P.
XX 15-FEB-2002; 2002US-0357253P.
XX 15-FEB-2002; 2002US-0357600P.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI Lioubin MN;
XX
XX WPI; 2003-156840/15.
XX
XX PT Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
XX acid, with a test agent.
XX
XX PS Disclosure; Page 47-51; 278pp; English.
XX
XX The present invention relates to a method for identifying candidate p53
XX pathway modulating agents. The method involves contacting an assay system
XX comprising purified chromatin organisation modifier (chromo) domain
XX helicase DNA binding proteins (CHD), nucleic acids, their functionally
XX active fragments or derivatives, with a test agent under conditions
XX where, but for the presence of the test agent, the system provides a
XX reference activity. The methods are useful for identifying modulators of
XX the p53 pathway as therapeutic targets for disorders associated with
XX defective p53 function, such as angiogenic disorders, apoptotic disorders
XX or cell proliferative disorders, e.g. cancer. The modulators are useful
XX as research reagents, diagnostics and therapeutics. The invention is also
XX useful in gene therapy. The present sequence is human CHD DNA
XX
XX SQ Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 10; Length 7764;
Best Local Similarity 87.5%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGCATGCTGTGACAC 25
Db 6509 AAAAACAAGCAGACTTTGACAC 6486

RESULT 20
ADK60458/c
ID ADK60458 standard; DNA; 7764 BP.

XX
XX ADK60458;
AC
XX
XX 06-MAY-2004 (first entry)
DE Angiogenesis differentially expressed gene GS-N32.
XX
XX
XX ds; vasotropic; antirheumatic; antiarthritic; hypotensive; antianginal;
KW antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;
KW angiogenesis; endothelial cell; diagnosis; tumor vascularization;
KW retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
KW ovary hyperstimulation; psoriasis; endometriosis; restenosis;
KW angioplasty; cicatrization; peripheral vascular disease; hypertension;
KW vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
KW ischemia; angina; myocardial infarction; chronic heart disease;
KW cardiac congestion; macular degeneration; osteoporosis.
XX
XX Homo sapiens.
XX
XX FR2836687-A1.
XX
XX 05-SEP-2003.
XX
XX 11-APR-2002; 2002FR-00004546.
XX
XX 04-MAR-2002; 2002FR-00002717.
XX
XX (GENE-) GENE SIGNAL.
XX PA (ALMA/) AL MAHMOOD S.
XX
XX PI Colin S, Schneider C, Al Mahmood S;
XX
XX WPI; 2004-013912/02.
XX P-PSDB; ADK60208.
XX
XX PT Compositions for diagnosing, prognosing and treating angiogenic disorders
PT including tumor vascularization and heart disease, comprise nucleic acid
PT or polypeptide differentially expressed in angiogenesis.
XX
XX Claim 2; SEQ ID NO 33; 424pp; French.
XX
XX The invention relates to a novel pharmaceutical composition active on
XX angiogenesis comprising an endothelial cell nucleic acid whose expression
XX is induced by an angiogenic factor and inhibited by an angiostatic agent
XX or its complement or fragment, a polypeptide sequence encoded by the
XX nucleic acid or its fragment, a molecule capable of inhibiting expression
XX of the nucleic acid or a molecule which binds to the polypeptide
XX sequence. The invention is used to diagnose, prognose or treat an
XX angiogenic disorder in a mammal, particularly a human. The disorder is
XX particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
XX Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
XX endometriosis associated with neovascularization, restenosis due to
XX angioplasty, overproduction of tissue due to cicatrization, a peripheral
XX vascular disease, hypertension, vascular inflammation, Raynaud disease,
XX aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,
XX myocardial infarction, chronic heart disease, cardiac congestion or
XX macular degeneration due to age or osteoporosis. This sequence
XX corresponds to a differentially expressed DNA used in the composition of
XX the invention.
XX
XX SQ Sequence 7764 BP; 2342 A; 1535 C; 1898 G; 1989 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 12; Length 7764;
Best Local Similarity 87.5%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGCATGCTGTGACAC 25
Db 6509 AAAAACAAGCAGACTTTGACAC 6486

RESULT 21
ADK60759/c

ID ADK60759 standard; DNA; 7764 BP.
 XX
 AC ADK60759;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Angiogenesis differentially expressed gene GS-N32.
 XX
 de; vasotropic; antirheumatic; antiarthritic; hypotensive; antianginal;
 KM antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;
 KM angiogenesis; endothelial cell; diagnosis; tumor vascularization;
 KM retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;
 KM ovary hyperstimulation; psoriasis; endometriosis; restenosis;
 KM angioplasty; cicatrization; peripheral vascular disease; hypertension;
 KM vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;
 KM ischemia; angina; myocardial infarction; chronic heart disease;
 KM cardiac congestion; macular degeneration; osteoporosis.
 XX
 OS Homo sapiens.
 XX
 PN FR283686-A1.
 XX
 PD 05-SEP-2003.
 XX
 PF 04-MAR-2002; 2002FR-00002717.
 XX
 PR 04-MAR-2002; 2002FR-00002717.
 XX
 PA (GENE-) GENE SIGNAL.
 XX (ALMA/) AL MAHMOOD S.
 PI Colin S, Schneider C, Al Mahmood S;
 DR WPI; 2004-013911/02.
 DR P-PSDB; ADK60509.
 XX
 PT Compositions containing nucleic acid or polypeptide differentially
 PT expressed in angiogenesis are useful to diagnose, prognosis and treat
 PT angiogenic disorders including tumor vascularization and heart disease.
 XX
 PS Claim 2; SEQ ID NO 33; 405bp; French.
 CC The invention relates to a novel pharmaceutical composition active on
 CC angiogenesis comprising an endothelial cell nucleic acid whose expression
 CC is induced by an angiogenic factor and inhibited by an angiostatic agent
 CC or its complement or fragment, a polypeptide sequence encoded by the
 CC nucleic acid or its fragment, a molecule capable of inhibiting expression
 CC of the nucleic acid or a molecule which binds to the polypeptide
 CC sequence. The invention is used to diagnose, prognosis or treat an
 CC angiogenic disorder in a mammal, particularly a human. The disorder is
 CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,
 CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,
 CC endometriosis associated with neovascularization, restenosis due to
 CC angioplasty, overproduction of tissue due to cicatrization, a peripheral
 CC vascular disease, hypertension, vascular inflammation, Raynaud disease,
 CC aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,
 CC myocardial infarction, chronic heart disease, cardiac congestion or
 CC macular degeneration due to age or osteoporosis. This sequence
 CC corresponds to a differentially expressed DNA used in the composition of
 CC the invention.
 XX
 SQ Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;
 XX
 Query Match 76.8%; Score 19.2; DB 12; Length 7764;
 Best local Similarity 87.5%; Pred. No. 5.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 AAAAAAAAAAGCATGCTGACAC 25
 DB 6509 AAAAAAGACGACGACTTGACAC 6486

ADP73081/C
 ID ADP73081 standard; DNA; 7764 BP.
 XX
 AC ADP73081;
 XX
 DT 12-AUG-2004 (first entry)
 XX
 DE Angiogenesis inhibitor human DNA sequence, GS-N32.
 XX
 KM inhibitor; angiogenesis; antisense nucleic acid; immunisation;
 KM angiogenic disorder; antiangiogenic; angiogenesis stimulator; cycostatic;
 KM dermatological; antiarthritic; antirheumatic; antiinflammatory;
 KM vasotropic; hypotensive; ophthalmological; antipsoriatic; cardiac;
 KM gene therapy; antisense gene therapy; tumor vascularization;
 KM retinopathies; rheumatoid arthritis; Crohn's disease; atherosclerosis;
 KM ovarian hyperstimulation; psoriasis; endometriosis; restenosis;
 KM tissue granulation; peripheral vascular disorder; hypertension;
 KM vascular inflammation; Raynaud's disease; aneurism; arterial restenosis;
 KM thrombophlebitis; lymphadenopathy; lymphedema; ischemia; angina;
 KM myocardial infarction; chronic heart disease; congestive heart disease;
 KM macular degeneration; human; de; gene.
 XX
 OS Homo sapiens.
 XX
 PN FR2843753-A1.
 XX
 PD 27-FEB-2004.
 XX
 PF 20-JUN-2003; 2003FR-00007507.
 XX
 PR 04-MAR-2002; 2002FR-00002717.
 XX
 PA (GENE/) GENE S.
 XX (ALMS/) AL M S.
 PI Colin S, Schneider C, Al MS;
 DR WPI; 2004-216677/21.
 DR P-PSDB; ADP73132.
 XX
 PT Antisense nucleic molecule useful as inhibitor of angiogenesis in the
 PT treatment of angiogenic disorders, e.g., rheumatoid arthritis,
 PT atherosclerosis and endometriosis.
 XX
 PS Claim 1; SEQ ID NO 33; 404bp; French.
 CC The invention relates to a novel inhibitor of angiogenesis comprising an
 CC active substance chosen from at least one of a nucleic acid molecule, an
 CC antisense nucleic acid molecule, a polypeptide or an antibody. The
 CC invention further comprises: an antisense nucleic acid sequence chosen
 CC from any of the sequences provided in the specification; preparation of
 CC the antibody comprising in vivo or in vitro immunisation of an
 CC immunocompetent animal cell, preferably of a vertebrate and most
 CC preferably of a mammal, with at least one of the polypeptide sequences
 CC chosen from a sequence provided in the specification; a mammalian
 CC expression vector comprising at least one antisense sequence chosen from
 CC an antisense nucleic acid provided in the specification; preparation of a
 CC genetically modified cell, that over- or under-expresses a gene
 CC implicated in an angiogenic disorder, comprising inserting the vector
 CC from above into a mammalian cell; a genetically modified cell that over-
 CC expresses or under-expresses at least one gene involved in angiogenesis
 CC by a nucleic acid sequence chosen from any of ADP73049 to ADP73338, as in
 CC the antisense nucleic acid selection of above, or a fragment of any of
 CC these; preparation of a cell line that stably expresses an expression
 CC vector. The angiogenesis inhibitor has the following activities:
 CC antiangiogenic; angiogenesis stimulator; angiogenesis inhibitor,
 CC cycostatic, dermatological, antiarthritic, antirheumatic,
 CC antiinflammatory, vasotropic, hypotensive, ophthalmological,
 CC antipsoriatic, and cardiac. The novel nucleic acid molecules of the
 CC invention may be used to treat disorders in gene therapy and antisense
 CC gene therapy. The nucleic acid sequences, proteins and antibodies as part
 CC of the therapeutic compositions are useful in treating a disorder of
 CC angiogenesis chosen from: tumor vascularization, retinopathies,

CC rheumatoid arthritis, Crohn's disease, atherosclerosis, ovarian
CC hyperstimulation, psoriasis, endometriosis associated with
CC neovascularization, restenosis due to balloon angioplasty, tissue
CC granulation due to scarification, peripheral vascular disorders,
CC hypertension, vascular inflammation, Raynaud's disease, aneurism,
CC arterial restenosis, thrombophlebitis, lymphadenopathy, lymphedema,
CC ischaemia, angina, myocardial infarction, chronic heart disease,
CC congestive heart disease, macular degeneration linked to age and
CC osteoporosis. This polynucleotide represents a human DNA sequence for the
CC creation of an angiogenesis inhibitor of the invention.
XX
SQ Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 12; Length 7764;
Best Local Similarity 87.5%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGCATGCTGTGACAC 25
DB 6509 AAAAAACAAGCAGCAGCTTTGACAC 6486
RESULT 23
ID ADX44407/c standard; DNA; 7764 BP.
XX
AC ADX44407;
XX
DT 19-MAY-2005 (first entry)
XX
DE Human chromodomain helicase DNA binding protein 2 gene SEQ ID NO:27.
XX
KW ds; gene; chromodomain helicase DNA binding protein 2; prognosis;
KW colorectal tumor; cytostatic; gastrointestinal disease; neoplasm.
XX
OS Homo sapiens.
XX
PN US2005048526-A1.
XX
PD 03-MAR-2005.
XX
PF 18-FEB-2004; 2004US-00782413.
XX
PR 27-AUG-2003; 2003US-00651237.
XX
PA (WANG/) WANG Y.
XX
PI Wang Y;
XX
DR WPI; 2005-212272/22.
XX
PT Assessing colorectal cancer status comprises identifying differential
PT modulation in a combination of genes or analyzing the gene expression
PT profiles of biological samples.
XX
PS Claim 7; SEQ ID NO 27; 114pp; English.
XX
XX The invention relates to a novel method for assessing colorectal cancer
CC status comprising identifying differential modulation in a combination of
CC genes selected from any of the 7 nucleotide sequences having 1221-1974 bp
CC fully defined in the specification (ADX44387-ADX44393). The method of the
CC invention has cytostatic activity, and may have a use in gene therapy.
CC The methods and kit of the invention are useful for assessing, diagnosing
CC or treating colorectal cancer. The present sequence represents a gene
CC used in the invention.
XX
SQ Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 14; Length 7764;
Best Local Similarity 87.5%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGCATGCTGTGACAC 25

DB 6509 AAAAAACAAGCAGCAGCTTTGACAC 6486
RESULT 24
ID ADY39318/c standard; DNA; 7764 BP.
XX
AC ADY39318;
XX
DT 19-MAY-2005 (first entry)
XX
DE Human CHD2 gene sequence - SEQ ID 27.
XX
KW colorectal cancer; CHD2; gene; ds.
XX
OS Homo sapiens.
XX
PN US2005048494-A1.
XX
PD 03-MAR-2005.
XX
PF 27-AUG-2003; 2003US-00651237.
XX
PR 27-AUG-2003; 2003US-00651237.
XX
PA (WANG/) WANG Y.
XX
PI Wang Y;
XX
DR WPI; 2005-212262/22.
XX
PT Assessing colorectal cancer status comprises identifying differential
PT modulation in a combination of genes or analyzing the gene expression
PT profiles of biological samples.
XX
PS Claim 7; SEQ ID NO 27; 114pp; English.
XX
XX The invention comprises a method of assessing colorectal cancer status,
CC the method involves identifying differential modulation in a combination
CC of genes. The method of the invention is useful for assessing, diagnosing
CC and treating colorectal cancer. The present DNA sequence represents a
CC human colorectal cancer-related gene that was used in an example of the
CC invention.
XX
SQ Sequence 7764 BP; 2341 A; 1536 C; 1898 G; 1989 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 14; Length 7764;
Best Local Similarity 87.5%; Pred. No. 5.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGCATGCTGTGACAC 25
DB 6509 AAAAAACAAGCAGCAGCTTTGACAC 6486
RESULT 25
ID AAK80953 standard; DNA; 17757 BP.
XX
AC AAK80953;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35765.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN MO200157182-A2.
XX

PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
PF
XX 31-JAN-2000; 2000US-01790658P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0198774P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225799P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236378P.

PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX MPI, 2001-483426/52.
DR

```
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 35765; 3071bp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 17757 BP; 5365 A; 2938 C; 3623 G; 5831 T; 0 U; 0 Other;
XX
Query Match 76.8%; Score 19.2; DB 4; Length 17757;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 8903 AAAAAAAAAAAGACTGTGACA 8926
XX
RESULT 26
ADD48759
ID ADD48759 standard; DNA; 28564 BP.
XX
AC ADD48759;
XX
DT 02-DEC-2004 (revised)
DT 29-JAN-2004 (first entry)
XX
DE Human gene AC004854, SEQ ID NO 14469.
XX
KW Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
OS Unidentified.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PE 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-033347P.
XX
PA (GEHO ) GEN HOSPITAL CORP.
PA (FARB ) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M,
XX
DR WPI; 2003-268312/26.
DR GENBANK; AC004854.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
```

```
XX Example 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (described in Table 3 of
CC the specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 28564 BP; 7803 A; 6729 C; 6423 G; 7609 T; 0 U; 0 Other;
XX
Query Match 76.8%; Score 19.2; DB 10; Length 28564;
Best Local Similarity 87.5%; Pred. No. 6.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 4076 AAAAAAAAAAAGATGACTTGTGACA 4099
XX
RESULT 27
AAC35631/C
ID AAC35631 standard; DNA; 579 BP.
XX
AC AAC35631;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10874.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
```

PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-01300449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134321P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0149930P.
PR 26-AUG-1999; 99US-0150566P.
PR 27-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151067P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.

PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159322P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161997P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 75.2%; Score 18.8; DB 3; Length 579;
Best Local Similarity 90.9%; Pred. No. 6.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AAAAAAAAAAGCATGCTGACAC 25
Db 520 AAAAAAAAAAGCATGCTGACAC 499

RESULT 28
ADX54464/C
ID ADX54464 standard; cDNA; 1659 BP.

XX AC ADX54464;

XX DT 21-APR-2005 (first entry)

XX DE Plant full length insert polynucleotide seqid 29204.

XX KM plant protectant; plant growth regulant; gene therapy; plant;

XX KM recombinant DNA construct; physical array; plant breeding marker;
cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;

XX KM extreme osmotic condition; pathogen tolerance; pest tolerance;
growth rate; cell cycle pathway; disease resistance;

XX KM galactomannan production; lignin production; plant growth regulator;
yield; plant growth; plant development; seed oil; protein yield;

XX KM protein content; gene; ss.

XX OS Unidentified.

XX OS US2004034888-A1.

XX PD 19-FEB-2004.

XX PF 28-APR-2003; 2003US-00425114.

XX PR 06-MAY-1999; 99US-00304517.

XX PR 05-NOV-2001; 2001US-00985678.

XX (LITU/) LITU J.

XX (ZHOU/) ZHOU Y.

XX (KOVA/) KOVALIC D K.

XX (SCRE/) SCREEN S E.

XX (TABAS/) TABASKA J E.

XX (CAOV/) CAO Y.

PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
DR New recombinant DNA construct, useful for improving plant tolerance to
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX Claim 1; SEQ ID NO 29204; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?docid:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.

XX SQ Sequence 1659 BP; 503 A; 297 C; 356 G; 503 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 13; Length 1659;
Best Local Similarity 90.9%; Pred. No. 6.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGCTGCA 22
Db 1384 AAAAAAAAAAGCATGCTGCA 1363

RESULT 29

ABLI7270
ID ABLI7270 standard; DNA; 2613 BP.

XX AC ABLI7270;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 3283.

XX KM Drosophila; developmental biology; cell signalling; insecticide;

XX KM pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX OS WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers BW;

XX WPI; 2001-656860/75.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.
 XX Claim 1; SEQ ID NO 3283; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB57737-
 CC AB87072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2613 BP; 874 A; 462 C; 460 G; 817 T; 0 U; 0 Other;
 XX
 Query Match 75.2%; Score 18.8; DB 4; Length 2613;
 Best Local Similarity 90.9%; Pred. No. 7.1e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGA 22
 DB 2488 AAAAAAAAAAGCATGACTGTGA 2509
 XX
 RESULT 30
 ABD33532/c
 XX ID ABD33532 standard; DNA; 44063 BP.
 XX
 AC ABD33532;
 DT 18-NOV-2004 (first entry)
 XX
 DE Human cancer-associated (CA) gene HD07-105.
 XX
 KM Human; Cancer-associated protein; CAP; cancer-associated gene; CA; gene;
 KM de; cancer; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO2004058146-A2.
 XX
 PD 15-JUL-2004.
 XX
 PF 15-DEC-2003; 2003WO-US040081.
 XX
 PR 17-DEC-2002; 2002US-00322281.
 XX
 PA (SAGE-) SAGES DISCOVERY INC.
 XX
 PI Morris DW, Malandro MS;
 XX
 DR WPI; 2004-499109/47.
 XX
 PT Novel human cancer associated protein encoded within open reading frame
 PT of cancer associated gene, useful as targets for diagnosing cancer.
 XX
 PS Claim 16; SEQ ID NO 718; 182pp; English.
 XX
 CC The invention relates to cancer-associated proteins (CAP) and the cancer-
 CC associated (CA) nucleic acids encoding them. The invention also relates
 CC to a method for treating cancers involving administering to a patient an
 CC inhibitor of CAP, and a method of screening for anticancer activity in a
 CC potential drug involving providing a cell that expresses a CA gene,
 CC contacting a tissue sample derived from a cancer cell with an anticancer
 CC drug candidate and monitoring the effect of the anticancer drug candidate
 CC on expression of the CA gene. The CAP proteins are useful for detecting
 CC cancer associated with expression of a CAP protein in a test cell sample
 CC and for screening for a bioactive agent capable of modulating the
 CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
 CC cancer, involving determining the expression of a CA nucleic acid in a
 CC tissue. This sequence represents a human CA gene of the invention. Note:

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 44063 BP; 11420 A; 9592 C; 9911 G; 12873 T; 0 U; 267 Other;
 XX
 Query Match 75.2%; Score 18.8; DB 13; Length 44063;
 Best Local Similarity 90.9%; Pred. No. 9.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGA 22
 DB 41585 AAAAAAAAAAGCATGACTGTGA 41564
 XX
 RESULT 31
 ABQ76617
 XX ID ABQ76617 standard; cDNA; 81440 BP.
 XX
 AC ABQ76617;
 DT 21-NOV-2002 (first entry)
 XX
 DE C. albicans BAX-associated cDNA fragment SEQ ID 659.
 XX
 KM Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
 KM vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
 KM apoptosis; fungal; yeast; infection; autoimmune disease; ischemia;
 KM neurodegeneration; cell death; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200264766-A2.
 XX
 PD 22-AUG-2002.
 XX
 PF 21-DEC-2001; 2001WO-EP015398.
 XX
 PR 22-DEC-2000; 2000EP-00870318.
 PR 04-JAN-2001; 2001EP-00870002.
 PR 09-JAN-2001; 2001EP-00870003.
 XX
 PA (JANC) JANSSEN PHARM NV.
 XX
 PI Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
 XX
 DR WPI; 2002-667002/71.
 DR P-PSDB; ABG93351.
 XX
 PT New isolated nucleic acid representing a synthetic BAX-gene, useful as
 PT medicament for treating, preventing and/or alleviating yeast or fungal
 PT infections or proliferative disorders, or for preventing apoptosis in
 PT certain diseases.
 XX
 PS Claim 36; Fig 2; 34pp; English.
 XX
 CC This invention describes a novel nucleic acid representing a synthetic
 CC Bax gene. The Bax gene of the invention is useful for identifying Bax-
 CC resistant yeast or fungi, identifying, or obtaining and identifying
 CC Candida spp. sequences that are differentially expressed in a pathway
 CC eventually leading to programmed cell death or identifying inhibitors or
 CC inhibitor sequences of Bax-induced cell death. The products of the
 CC invention have cytostatic, fungicide; immunosuppressive, virucide and
 CC vasotropic activity and can be used in vaccines or for gene therapy. The
 CC isolated nucleic acids, polypeptides, pharmaceutical compositions,
 CC antisense molecules and antibodies are useful as medicaments or in
 CC preparing a medicament for treating, preventing and/or alleviating
 CC diseases associated with yeast or fungi or proliferative disorders, such
 CC as cancer, or for preventing apoptosis in certain diseases. The compounds
 CC or polypeptides, or the genetically modified organism are useful for
 CC preparing a medicament for modifying the endogenous flora of humans and
 CC other mammals. The vaccine is useful for immunising against yeast or
 CC fungal infections. Apoptosis-related diseases include autoimmune disease,

CC ischaemia, diseases related with viral infections or neurodegenerations.
CC This sequence represents a polynucleotide associated with the Bax gene
CC described in the disclosure of the invention
XX
SQ Sequence 81440 BP; 28347 A; 12937 C; 12536 G; 27617 T; 0 U; 3 Other;
Query Match 75.2%; Score 18.8; DB 6; Length 81440;
Best Local Similarity 90.9%; Pred. No. 1e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CY 1 AAAAAAAAAAGCATGACTGTGA 22
Db 486 AAAAAAAAAATATCATGCTGTA 507
RESULT 32
AA161373_2
Continuation (3 of 6) of AA161373 from base 200001 (Soybean 318013 region A3, SEQ ID NO:
WP Sequence split into 6 fragments LOCUS AA161373 Accession AA161373
WP Fragment Name Begin End
WP AA161373_0 1 110000
WP AA161373_1 100001 210000
WP AA161373_2 200001 310000
WP AA161373_3 300001 410000
WP AA161373_4 400001 510000
WP AA161373_5 500001 513445
Query Match 75.2%; Score 18.8; DB 5; Length 110000;
Best Local Similarity 90.9%; Pred. No. 1e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CY 1 AAAAAAAAAAGCATGACTGTGA 22
Db 29055 AAAAAAAAAAGCATGCTGTTA 29076
RESULT 33
ABV01013/c
ID ABV01013 standard; cDNA; 361 BP.
XX
AC ABV01013;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 1004.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.

XX
PS Claim 1; Page 257; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC / (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 361 BP; 102 A; 86 C; 72 G; 99 T; 0 U; 2 Other;
Query Match 74.4%; Score 18.6; DB 5; Length 361;
Best Local Similarity 84.0%; Pred. No. 7.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
CY 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 212 AAAAAAAAAAGCACCTGCCACAC 188
RESULT 34
ABV10182/c
ID ABV10182 standard; cDNA; 396 BP.
XX
AC ABV10182;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 10173.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 1632; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 396 BP; 120 A; 90 C; 72 G; 114 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 5; Length 396;
 Best Local Similarity 84.0%; Pred. No. 7.1e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
 DB 195 AAAAAAAAAAGCACCTGCCACAC 171
 RESULT 35
 ABV31353/c
 ID ABV31353 standard; cDNA; 415 BP.
 XX
 AC ABV31353;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker CDNA 31344.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JB;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 Claim 1; Page 6751-6752; 11750pp; English.
 XX
 The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement; (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 415 BP; 121 A; 95 C; 84 G; 115 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 5; Length 415;
 Best Local Similarity 84.0%; Pred. No. 7.2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
 DB 233 AAAAAAAAAAGCACCTGCCACAC 209
 RESULT 36
 ABV40322/c
 ID ABV40322 standard; cDNA; 415 BP.
 XX
 AC ABV40322;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker CDNA 40313.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JB;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 Claim 1; Page 8141; 11750pp; English.
 XX
 The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement; (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 415 BP; 121 A; 95 C; 84 G; 115 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 5; Length 415;
 Best Local Similarity 84.0%; Pred. No. 7.2e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
 DB 233 AAAAAAAAAAGCACCTGCCACAC 209

```
RESULT 37
ABV43675/c
ID ABV43675 standard; cDNA; 415 BP.
XX
AC ABV43675;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 43666.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200106860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001MO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 8693; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 415 BP; 121 A; 95 C; 84 G; 115 T; 0 U; 0 Other;
XX
Query Match 74.4%; Score 18.6; DB 5; Length 415;
Best Local Similarity 84.0%; Pred. No. 7.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 233 AAAAAAAAAAGCATGACTGTGACAC 209
RESULT 38
ACN59204/c
ID ACN59204 standard; cDNA; 561 BP.
XX
AC ACN59204;
XX
DT 02-DEC-2004 (first entry)
XX
DE Cotton gymnocium tissue EST Clone ID: LIB3829-023-Q6-K6-E5, SEQ:13985.
```

```
XX
KM Cotton; plant; EST; expressed sequence tag; transgenic plant; gymnocium;
KM variety Nucotton33B; library LIB3829; molecular tag; molecular marker;
KM genetic mapping; molecular mapping; seed germination; plant growth;
KM plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS Gossypium hirsutum.
XX
PN US2004123340-A1.
XX
PD 24-JUN-2004.
XX
PE 12-DEC-2001; 2001US-00021323.
XX
PR 14-DEC-2000; 2000US-0255619P.
XX
PA (DEIK/) DEIKMAN J.
PA (FENG/) FENG P C C.
PA (FINC/) FINCHER K L.
PA (ZIEG/) ZIEGLER T E.
XX
PI Deikman J, Feng PCC, Fincher KL, Ziegler TE;
XX
DR WPI; 2004-479808/45.
XX
PT New isolated nucleic acid molecule that encodes a plant protein or its
PT fragment, useful for isolating a variety of agronomically significant
PT genes associated with plant growth, quality or yield, and as molecular
PT tags to map genes.
XX
PS Claim 1; SEQ ID NO 13985; 34pp; English.
XX
CC The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC from primed or non-primed seeds from variety DP50B, mature seeds from
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gymnocium
CC tissue, developing fibres, carpel walls and septa from variety
CC Nucotton33B. The invention also relates to substantially purified
CC proteins or their fragments encoded by nucleic acid molecules of the
CC invention, and to transformed plants having a nucleic acid construct
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC molecular tags to isolate genetic regions, to isolate genes, to map
CC genes, to determine gene function and to determine whether genes are
CC members of a particular gene family. The nucleic acid molecules may be
CC used for isolating a variety of agronomically significant genes
CC associated with plant growth, quality, yield, and could also serve as
CC links in metabolic and catabolic pathways. The nucleic acid molecules are
CC also useful for identifying genes important in initiating and maintaining
CC seed germination or that may be used to mitigate stresses encountered
CC during seed germination. The ESTs additionally enable the acquisition of
CC promoters and cis-regulatory elements which will be useful to express
CC agronomically significant genes in these tissues and/or other tissues,
CC and also permits the acquisition of molecular markers useful in breeding
CC schemes, genetic and molecular mapping, and in cloning of agronomically
CC significant genes. The nucleic acid molecules are further useful for
CC detecting the expression level or pattern of a protein or mRNA and for
CC detecting the presence or quantity of a protein by tissue printing. The
CC present sequence represents a specifically claimed EST isolated from a
CC cotton variety Nucotton33B gymnocium tissue cDNA library (LIB3829). The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the US
CC patent office at seqdata.uspto.gov/sequence.htm?DocID=US20040123340
XX
SQ Sequence 561 BP; 123 A; 99 C; 171 G; 168 T; 0 U; 0 Other;
XX
Query Match 74.4%; Score 18.6; DB 13; Length 561;
Best Local Similarity 84.0%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 150 AAAAAAAAAAGCATGACTGTGACAC 126
```

```

RESULT 39
ABV26507
ID ABV26507 standard; cDNA; 686 BP.
XX
XX
AC ABV26507;
XX
XX 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 26498.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
XX 23-AUG-2001.
XX
PD 20-FEB-2001; 2001WO-US005171.
XX
PF 17-FEB-2000; 2000US-0183119P.
XX
PR 16-MAR-2000; 2000US-0189862P.
XX
PR 25-MAY-2000; 2000US-0207454P.
XX
PR 09-JUN-2000; 2000US-0211314P.
XX
PR 18-JUL-2000; 2000US-0219007P.
XX
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
XX MPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 5355; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 686 BP; 208 A; 126 C; 170 G; 179 T; 0 U; 3 Other;
XX
Query Match 74.4%; Score 18.6; DB 5; Length 686;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 501 AAAAAAAAAAGCACCTGCCACAC 525

```

```

DE Human prostate expression marker cDNA 20656.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183119P.
XX
XX 16-MAR-2000; 2000US-0189862P.
XX
XX 25-MAY-2000; 2000US-0207454P.
XX
XX 09-JUN-2000; 2000US-0211314P.
XX
XX 18-JUL-2000; 2000US-0219007P.
XX
XX 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JB;
XX
XX MPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 3388; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 686 BP; 208 A; 126 C; 170 G; 179 T; 0 U; 3 Other;
XX
Query Match 74.4%; Score 18.6; DB 5; Length 686;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 501 AAAAAAAAAAGCACCTGCCACAC 525

```

```

RESULT 40
ABV20665
ID ABV20665 standard; cDNA; 686 BP.
XX
XX AC ABV20665;
XX
XX 13-SEP-2002 (first entry)
XX

```

```

RESULT 41
ABV22901
ID ABV22901 standard; cDNA; 686 BP.
XX
XX AC ABV22901;
XX
XX 13-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 22892.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX

```

PD 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US0005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
PI WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 4057; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC / (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 686 BP; 208 A; 126 C; 170 G; 179 T; 0 U; 3 Other;
SQ
Query Match 74.4%; Score 18.6; DB 5; Length 686;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGACAC 25
DB 501 AAAAAAAAAAGCATGCTGACAC 525
RESULT 42
ABV28735
ID ABV28735 standard; cDNA; 686 BP.
XX
XX AC ABV28735;
XX
XX DT 16-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 28726.
XX
XX KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200160860-A2.
XX
XX PD 23-AUG-2001.
XX
XX PF 20-FEB-2001; 2001WO-US0005171.
XX
XX PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR

PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
PI WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 6034; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC / (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 686 BP; 208 A; 126 C; 170 G; 179 T; 0 U; 3 Other;
SQ
Query Match 74.4%; Score 18.6; DB 5; Length 686;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGACAC 25
DB 501 AAAAAAAAAAGCATGCTGACAC 525
RESULT 43
ADQ25229
ID ADQ25229 standard; DNA; 711 BP.
XX
XX AC ADQ25229;
XX
XX DT 26-AUG-2004 (first entry)
XX
XX DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 8049.
XX
XX KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM de.
XX
XX OS Homo sapiens.
XX
XX PN WO2004048938-A2.
XX
XX PD 10-JUN-2004.
XX
XX PF 26-NOV-2003; 2003WO-US038193.
XX
XX PR 26-NOV-2002; 2002US-0429739P.
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
XX
XX PA Atiz N, Ginsburg WM, Zlotnik A;
XX
XX PI WPI; 2004-441208/41.
XX
XX DR Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.
XX
XX Example 2; SEQ ID NO 8049; 210pp; English.
PS

XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 711 BP; 187 A; 166 C; 196 G; 162 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 12; Length 711;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0

1 AAAAAAAAAAGCATGCTGTGCAC 25
||||| ||||| |||||
645 AAAAAAAAAAAGTACTGCGAC 669

RESULT 44
ABN69451/C
ID ABN69451 standard; DNA; 1194 BP.
XX
XX ABN69451;
AC
XX
XX 01-JUL-2002 (first entry)
DT
XX
XX Streptococcus polymuclotide SEQ ID NO 6815.
DE
XX
XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus pyogenes.
OS
XX
XX MO200234771-A2.
PN
XX
XX 02-MAY-2002.
PD
XX
XX 29-OCT-2001; 2001MO-GB004789.
PF
XX
XX 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
PPA
XX
XX Telford J, Maignani V, Margart Y Rosi, Grandi G, Fraser C;
PI Tetteijn H;
XX
XX MPI: 2002-352536/38.
DR P-PSDB; ABP28820.
XX
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX
XX Claim 7; Page 3844; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B
Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
the specification. The proteins have antibacterial and anti-inflammatory
activity. (I), nucleic acids encoding (II), ABN6044-ABN71526 and
antibodies that bind (I) are used in the manufacture of medicaments for

CC	The treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
XX	Streptococcus proteins
SQ	Sequence 1194 BP; 328 A; 185 C; 247 G; 434 T; 0 U; 0 Other;
OY	Query Match 74.4%; Score 18.6; DB 6; Length 1194; Best Local Similarity 84.0%; Pred.No.7.9e+02; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0.
Dn	1 AAAAAAAAAAGCATGACTGTGCAC 25 881 AAAAATAAAAGTATTACTGTGCAC 857
RESULT 45	
ID	ADX13182/c
XX	ADX13182 standard; cDNA; 1268 BP.
AC	ADX13182;
DT	21-APR-2005 (first entry)
DB	Plant full length insert polynucleotide seqid 7757.
XX	
KW	plant protectant; plant growth regulant; gene therapy; plant;
KW	recombinant DNA construct; physical array; plant breeding marker;
KW	cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW	extreme osmotic condition; pathogen tolerance; pest tolerance;
KW	growth rate; cell cycle pathway; disease resistance;
KW	galactomannan production; lignin production; plant growth regulator;
KW	yield; plant growth; plant development; seed oil; protein yield;
KW	protein content; gene; ss.
OS	Unidentified.
XX	
PN	US2004034888-A1.
XX	
PD	19-FEB-2004.
XX	
Pf	28-APR-2003; 2003US-00425114.
PR	06-MAY-1999; 99US-00304517.
XX	
PR	05-NOV-2001; 2001US-00385678.
XX	
PA	(LNUJ/) LNU J.
PA	(ZHOU/) ZHOU Y.
PA	(KOVA/) KOVALIC D K.
PA	(SCRE/) SCREEN S E.
PA	(TABAS/) TABASKA J E.
PA	(CAOY/) CAO Y.
XX	
PI	Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
DR	WPI; 2004-180133/17.
XX	
PT	New recombinant DNA construct, useful for improving plant tolerance to
PT	cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT	pests, for conferring increased resistance to plant disease, or for
PT	improving yield.
PS	Claim 1; SEQ ID NO 7757, 15pp; English.
XX	
CC	The invention describes a recombinant DNA construct comprising a
CC	polynucleotide consisting of a sequence encoding an amino acid sequence

CC available in electronic form from the US patent office at
CC ftp://seqdata.uspto.gov/sequence.html?docID:2004034688. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant diseases, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
CC
SQ Sequence 1268 BP; 377 A; 245 C; 292 G; 354 T; 0 U; 0 Other;
Query Match 74.4%; Score 18.6; DB 13; Length 1268;
Best Local Similarity 84.0%; Pred. No. 8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGCATGACTGTGACAC 25
DB 915 AAAAAAAAAAGAAAGACGATGACAC 891
RESULT 46
ADE61980
ID ADE61980 standard; DNA; 1612 BP.
AC ADE61980;
XX
XX 29-JAN-2004 (first entry)
DE Rat gene S85184, SEQ ID NO 7909.
XX
XX Rat; ds; gene; pain; neuronal tissue; gene therapy;
KM spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
XX WO2003016475-A2.
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
PI WPI; 2003-268312/26.
DR GENBANK; S85184.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a

CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which is differentially expressed during pain. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp://wipo.int/pub/publ/published_pct_sequences.
CC
SQ Sequence 1612 BP; 452 A; 295 C; 375 G; 490 T; 0 U; 0 Other;
Query Match 74.4%; Score 18.6; DB 10; Length 1612;
Best Local Similarity 84.0%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGCATGACTGTGACAC 25
DB 1471 AAAAAAAAAATTCATGATTTATAC 1495
RESULT 47
ADE61984
ID ADE61984 standard; DNA; 1612 BP.
AC ADE61984;
XX
XX 29-JAN-2004 (first entry)
DE Rat gene S85184, SEQ ID NO 7913.
XX
XX Rat; ds; gene; pain; neuronal tissue; gene therapy;
KM spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
XX WO2003016475-A2.
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
PI WPI; 2003-268312/26.
DR GENBANK; S85184.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotide or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
 CC specification) which is differentially expressed during pain. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1612 BP; 452 A; 295 C; 375 G; 490 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 10; Length 1612;

Best Local Similarity 84.0%; Pred. No. 8.2e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACAC 25

Db 1471 AAAAAAAAAATCATGATGTATAC 1495

RESULT 48

ADV97351

ID ADV97351 standard; DNA; 3216 BP.

XX ADV97351;

XX 24-FEB-2005 (first entry)

DE N tabacum histidine kinase-like 2 (NTHK2) gene.

XX NTHK2; plant; plant breeding; drought resistance;

KW Nicotiana tabacum histidine kinase-like 2; gene; ds.

XX Nicotiana tabacum.

XX CN1438241-A.

XX 27-AUG-2003.

XX 11-FEB-2002; 2002CN-00105035.

XX 11-FEB-2002; 2002CN-00105035.

XX (GENE-) GENETICS INST CHINESE ACADE SCI.

XX Zhang J, Chen S, Zhang Z;

XX WPI; 2004-157344/16.

XX P-PSDB; ADV97350.

PT Reverse-correlation-resetting signal transfer gene of plant and protein

XX coded thereof.

XX Claim 6; Page 3-5; 15pp; Chinese.

CC The invention discloses a signal-conveying gene NTHK2 in plants, the
 CC protein encoded by it and its characters of damage, drought and high-
 CC temperature guide in plants. The encoded protein and heterozygous
 CC ethylene receptor protein have higher isogeny but are not of the same
 CC kind, and the encoded protein is novel relative to contrary-resistance of
 CC plants. The invention also refers to their applications in plant gene of
 CC engineering. The present sequence is that of the N tabacum NTHK2 gene of
 CC the invention.

XX Sequence 3216 BP; 959 A; 539 C; 699 G; 1019 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 13; Length 3216;

Best Local Similarity 84.0%; Pred. No. 8.8e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACAC 25

Db 2902 AAAAAAAAAATCATGATGTATAC 2926

RESULT 49

AAH18707

ID AAH18707 standard; cDNA; 5020 BP.

XX AAH18707;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:18976.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 18976; 2537bp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH1629 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 5020 BP; 1170 A; 1222 C; 1217 G; 1411 T; 0 U; 0 Other;
Query Match 74.4%; Score 18.6; DB 4; Length 5020;
Best Local Similarity 84.0%; Pred. No. 9.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
CY 1 AAAAAAAAAAGCATGCTGACAC 25
Db 2499 AAAAAAAAAAGCATTTATGACAC 2523
RESULT 50
AAL06441
ID AAL06441 standard; DNA; 6495 BP.
XX
AC AAL06441;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9129.
XX
KM Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 16-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

```
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251909P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX DR WPI; 2001-465570/50.
XX
XX PT Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX PS Disclosure; SEQ ID NO 9129; 1297bp + Sequence Listing; English.
XX
XX CC The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
XX SQ Sequence 6495 BP; 2321 A; 1201 C; 1251 G; 1722 T; 0 U; 0 Other;
XX
XX Query Match 74.4%; Score 18.6; DB 4; Length 6495;
XX Best Local Similarity 84.0%; Pred. No. 9,4e+02;
XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
XX |||||
XX |||||
XX |||||
XX |||||
XX |||||
XX DB 2315 AAAAAAAAAAGAAAGGTGACAC 2339
```

Search completed: December 14, 2005, 02:42:49
Job time : 212.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:34:03 ; Search time 1752.1 Seconds
(without alignments)
667.586 Million cell updates/sec

Title: US-10-681-773-5

Perfect score: 25

Sequence: 1 aaaaaaaaaagcatgactgtgacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	537	6	CA397104 c886c01.y
C 2	23.4	93.6	933	5	BX456744 BX456744
C 3	23.4	93.6	1468	4	CR625326 BX156774
C 4	21.4	85.6	874	5	BUI31141 BUI31141
C 5	20.8	83.2	622	9	CE138790 CE138790
C 6	20.8	83.2	625	9	BUI38790 BUI38790
C 7	20.8	83.2	645	7	CK535043 CK535043
C 8	20.8	83.2	888	10	AG880409 AG880409
C 9	20.8	83.2	426	5	BU465605 BU465605
C 10	20.4	81.6	426	8	W11074 W11074
C 11	20.4	81.6	434	11	DE112709 DE112709
C 12	20.4	81.6	436	2	BG588669 BG588669
C 13	20.4	81.6	525	5	BX516673 BX516673
C 14	20.4	81.6	612	6	CB840280 CB840280
C 15	20.4	81.6	637	3	BP766673 BP766673
C 16	20.4	81.6	756	10	CL141446 CL141446
C 17	20.4	81.6	816	6	CA920541 CA920541
C 18	20.4	81.6	821	7	CK787550 CK787550
C 19	20.4	81.6	892	3	BI904008 BI904008
C 20	20.2	80.8	210	1	AA936662 AA936662
C 21	20.2	80.8	330	9	AQ097896 AQ097896
C 22	20.2	80.8	404	10	CW730653 CW730653

23	20.2	80.8	412	1	AA931550
24	20.2	80.8	446	10	CW730610
C 25	20.2	80.8	447	7	CK059387
C 26	20.2	80.8	510	6	CB641312
C 27	20.2	80.8	525	10	CW730743
C 28	20.2	80.8	556	8	CV844822
C 29	20.2	80.8	557	9	BH569468
C 30	20.2	80.8	578	7	CK072813
C 31	20.2	80.8	579	10	CW755031
C 32	20.2	80.8	582	3	BJ079563
C 33	20.2	80.8	611	9	AQ328946
C 34	20.2	80.8	653	10	CL420341
C 35	20.2	80.8	664	9	CE223739
C 36	20.2	80.8	716	2	BE212779
C 37	20.2	80.8	749	10	CM870653
C 38	20.2	80.8	752	9	AQ329214
C 39	20.2	80.8	760	6	CB671063
C 40	20.2	80.8	791	9	BH125511
C 41	20.2	80.8	796	10	BX163609
C 42	20.2	80.8	808	10	BX156386
C 43	20.2	80.8	854	10	CL762965
C 44	20.2	80.8	917	8	DR922466
C 45	20.2	80.8	950	11	CNS04VMB
C 46	20.2	80.8	1253	10	CG751092
C 47	20.2	80.8	351	10	CE777289
C 48	20.2	80.8	866	10	CNS02XVM
C 49	19.8	79.2	344	9	AZ576423
C 50	19.8	79.2	400	9	BH189388
C 51	19.8	79.2	400	11	CNS0776T
C 52	19.8	79.2	408	9	BZ891446
C 53	19.8	79.2	469	9	CE021612
C 54	19.8	79.2	503	8	DN429833
C 55	19.8	79.2	552	3	BQ130854
C 56	19.8	79.2	562	9	AO513411
C 57	19.8	79.2	582	9	CE006606
C 58	19.8	79.2	602	10	AG018806
C 59	19.8	79.2	634	9	B90334
C 60	19.8	79.2	638	5	BX509462
C 61	19.8	79.2	722	5	BU475819
C 62	19.8	79.2	750	10	AG444668
C 63	19.8	79.2	804	2	BF968713
C 64	19.8	79.2	835	10	AG848866
C 65	19.8	79.2	906	8	DN516976
C 66	19.8	79.2	966	7	CK187130
C 67	19.8	79.2	979	8	DN519682
C 68	19.8	79.2	1413	10	AG465923
C 69	19.4	77.6	350	1	AM681988
C 70	19.4	77.6	427	9	BZ18977
C 71	19.4	77.6	495	5	BX250833
C 72	19.4	77.6	497	11	CNS07PF78
C 73	19.4	77.6	535	5	BX516671
C 74	19.4	77.6	690	9	BH978184
C 75	19.4	77.6	701	6	CF385413
C 76	19.4	77.6	775	8	DR160322
C 77	19.4	77.6	787	8	DN096033
C 78	19.4	77.6	814	8	CV823258
C 79	19.4	77.6	956	10	CNS01PF9
C 80	19.4	77.6	1007	11	CNS04OSB
C 81	19.4	77.6	1209	5	BQ653410
C 82	19.2	76.8	193	9	AZ720219
C 83	19.2	76.8	200	1	AM890885
C 84	19.2	76.8	216	2	BB591303
C 85	19.2	76.8	238	7	CNS62907
C 86	19.2	76.8	255	1	AA416289
C 87	19.2	76.8	258	8	CV993135
C 88	19.2	76.8	283	9	AZ406588
C 89	19.2	76.8	310	1	AM805973
C 90	19.2	76.8	314	1	AW749265
C 91	19.2	76.8	322	10	CL880409
C 92	19.2	76.8	333	8	H75606
C 93	19.2	76.8	343	2	BF365769
C 94	19.2	76.8	351	2	BG514452
C 95	19.2	76.8	361	7	CK562766

```

96 19.2 76.8 374 10 CE467473
97 19.2 76.8 385 8 N91187
98 19.2 76.8 394 5 BM546642
99 19.2 76.8 405 1 BM546642
100 19.2 76.8 415 1 A1384286
101 19.2 76.8 436 11 CR153989
102 19.2 76.8 437 8 N91208
103 19.2 76.8 444 2 BE245184
104 19.2 76.8 447 1 A1384312
105 19.2 76.8 450 7 CV740296
106 19.2 76.8 457 8 CX535997
107 19.2 76.8 461 9 A0676439
108 19.2 76.8 469 1 A0655977
109 19.2 76.8 476 10 CL873239
110 19.2 76.8 484 9 A2518831
111 19.2 76.8 485 10 CL896423
112 19.2 76.8 490 8 CX535992
113 19.2 76.8 492 7 B36351
114 19.2 76.8 497 7 CO843898
115 19.2 76.8 499 2 BF707279
116 19.2 76.8 504 6 CB783761
117 19.2 76.8 506 9 CC875215
118 19.2 76.8 509 2 BF938005
119 19.2 76.8 512 1 AL380982
120 19.2 76.8 515 7 CV748613
121 19.2 76.8 519 9 AZ385283
122 19.2 76.8 519 9 CC593192
123 19.2 76.8 524 10 CZ803433
124 19.2 76.8 525 9 A0760854
125 19.2 76.8 525 10 CE511706
126 19.2 76.8 527 9 CE118656
127 19.2 76.8 532 10 CE804348
128 19.2 76.8 532 9 A0442684
129 19.2 76.8 536 7 CE681959
130 19.2 76.8 543 10 BX120656
131 19.2 76.8 547 9 AZ982190
132 19.2 76.8 551 9 A0786503
133 19.2 76.8 553 1 AM805979
134 19.2 76.8 555 6 CB479889
135 19.2 76.8 559 9 AZ332180
136 19.2 76.8 568 9 A0513402
137 19.2 76.8 570 9 CV748497
138 19.2 76.8 572 9 CE058929
139 19.2 76.8 573 10 CE519482
140 19.2 76.8 573 10 CL379905
141 19.2 76.8 574 11 CR326318
142 19.2 76.8 575 7 CV748082
143 19.2 76.8 585 7 CV294712
144 19.2 76.8 591 9 CE092606
145 19.2 76.8 594 7 CV748439
146 19.2 76.8 596 10 CL147798
147 19.2 76.8 596 10 CL139733
148 19.2 76.8 599 10 CM292923
149 19.2 76.8 603 10 CE662963
150 19.2 76.8 604 7 CV748493

```

ALIGNMENTS

```

RESULT 1
CA397104/c 537 bp mRNA linear EST 06-NOV-2002
LOCUS CA397104.1 GI:24734101
DEFINITION cs86c01.y2 Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unambiguated): cs Homo sapiens cDNA clone cs86c01
5', mRNA sequence.

```

```

ACCESSION CA397104
VERSION CA397104.1 GI:24734101
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

```

```

REFERENCE 1 (bases 1 to 537)
AUTHORS Mischow,G., Bernstein,S.D., Wyatt,M.K., Farris,R.N., Behal,A.,
Touhman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human RPE/choroid for the
NEI/Rank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
JOURNAL Mol. Vis. 8 (4), 205-220 (2002)
PUBMED 12107410
COMMENT Contact: Mischow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Place: 86 row: c column: 01
Seg primer: M13R1 reverse primer (ABI).
Location/Qualifiers
1. 537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="cs86c01"
/tissue_type="RPE/choroid"
/dev_stage="Adult"
/lab_host="EMDH108"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unambiguated): cs"
/note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
eyes (75-80 years old) yielded approximately 600 mg of
dissected RPE/choroid tissue. This in turn yielded 340 ug
of total RNA and 7 ug of mRNA. A directionally cloned cDNA
library in the pCMVSPORT6 vector was constructed at life
technologies (Rockville, MD; now part of Invitrogen Corp),
essentially following the protocols of the Superscript
Plasmid System (Invitrogen Corp).
<http://www.invitrogen.com/>. The library code
designation was cs. For this library, cDNA inserts were
cloned into the NotI/MluI sites of the vector. EST
analysis was performed on the unambiguated library at the
NIH Intramural Sequencing Center (NISC)."
```

ORIGIN

```

Query Match 100.0%; Score 25; DB 6; Length 537;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 113 AAAAAAAAAAGCATGACTGTGACAC 89

```

```

RESULT 2
BX456744/c 933 bp mRNA linear EST 06-MAY-2004
LOCUS BX456744 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP003YD18
DEFINITION 5-PRIME. mRNA sequence.
ACCESSION BX456744
VERSION BX456744.2 GI:47066661
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 933)
REFERENCE 1 (bases 1 to 933)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:11030820.
Contact: Genoscope
Genoscope - Centre National de Sequencage

```


genomic survey sequence.
 CE138790
 VERSION CE138790.1 GI:35246373
 KEYWORDS GSS.
 SOURCE
 ORGANISM *Canis familiaris* (dog)
Canis familiaris
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 622)
 Kirkness, E. F., Bafna, V., Halpern, A. L., Levy, S., Remington, K., Rusch, D. B., Delcher, A. L., Pop, M., Wang, W., Fraser, C. M. and Venter, J. C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 PUBMED 14512627
 CONTACT: Kirkness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirkness@tigr.org
 Class: shotgun.

FEATURES
 source
 Location/Qualifiers
 1..622
 /organism="Canis familiaris"
 /mol_type="genomic DNA"
 /strain="standard Poodle"
 /db_xref="taxon:9615"
 /clone_11b="Dog Library"
 /note="Site 1: BactX; Libraries were prepared from peripheral blood"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 9; Length 622;
 Best Local Similarity 91.7%; Pred. No. 9.2e+02; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
 |||||
 606 AAAAAAAAAAGCATGACTGTGACA 583

Db

RESULT 6
 BU199572/c 625 bp mRNA linear EST 25-NOV-2002
 LOCUS 60415561F1 CSEQCHN03 Gallus gallus cDNA clone CHEST1003e14 5',
 DEFINITION mRNA sequence.
 ACCESSION BU199572
 VERSION BU199572.1 GI:25362841
 KEYWORDS EST.
 SOURCE *Gallus gallus* (chicken)
 ORGANISM *Gallus gallus*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus.
 1 (bases 1 to 625)
 Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 JOURNAL 12445392
 PUBMED
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008950
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..625
 source

/organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, HiseX"
 /db_xref="taxon:9031"
 /clone="CHEST1003e14"
 /tissue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_11b="CSEQCHN03"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS 1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN
 Query Match 83.2%; Score 20.8; DB 5; Length 625;
 Best Local Similarity 91.7%; Pred. No. 9.2e+02; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGACTGTGACAC 25
 |||||
 314 AAAAAAAAAAGCATGACTGTGAC 291

Db

RESULT 7
 CK535043 645 bp mRNA linear EST 05-OCT-2004
 LOCUS tswb0_003395.y1 swg Bombyx mori cDNA, mRNA sequence.
 DEFINITION CK535043
 ACCESSION CK535043
 VERSION CK535043.1 GI:40919497
 KEYWORDS EST.
 SOURCE *Bombyx mori* (domestic silkworm)
 ORGANISM *Bombyx mori*
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Dictyoptera; Bombycoidea; Bombycidae; Bombyx.
 1 (bases 1 to 645)
 Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., Li, C., Shen, Y., Lan, X., Yuan, D., Li, T., Xu, H., Yang, G., Wan, Y., Zhu, Y., Yu, M., Shen, W., Wu, D., Xiang, Z., Yu, J., Wang, J., Li, R. Q., Shi, J. P., Li, H., Li, G. Y., Su, J. N., Wang, X. L., Li, Q. Q., Zhang, Z. J., Wu, Q. F., Li, J., Zhang, Q. P., Xu, J. Z., Sun, H. B., Dong, L., Liu, D. Y., Zhao, S. L., Zhao, X. L., Meng, Q. S., Lan, F. D., Huang, X. G., Li, Y. Z., Fang, F., Li, C. F., Li, D. M., Sun, Y. Q., Zhang, Z. P., Yang, Z., Huang, Y. Q., Xi, Y., Qi, Q. H., He, D. D., Huang, H. Y., Zhang, X. W., Wang, Z. Q., Li, W. J., Cao, Y. Z., Yu, Y., Yu, H., Ji, H., Ye, J., Chen, H., Zhou, Y., Liu, B., Wang, J., Ye, J., Ji, H., Li, S., Ni, P., Zhang, J., Zhang, Y., Zheng, H., Mao, B., Mao, B., Ye, C., Li, S., Wang, J., Wong, G. K. and Yang, H.
 A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*)
 Science 306 (5703), 1937-1940 (2004)
 JOURNAL 15591204
 PUBMED
 COMMENT Contact: Zhonghui Xiang
 Southwest Agricultural University
 Chongqing Beibei
 Tel: 86-23-68251123
 Fax: 86-23-68251128
 Email: xzh@swu.cn.
 Location/Qualifiers
 1..645
 /organism="Bombyx mori"
 source


```

/mot_type="mRNA"
/strain="Dazhao(P50)"
/db_xref="taxon:7091"
/sex="male"
/tissue_type="testis"
/dev_stage="5ch-instar day-3 larva"
/clone_lib="swg"
/notes="vector: pbluescript II SK(+)"

ORIGIN
Query Match      83.2%; Score 20.8; DB 7; Length 645;
Best Local Similarity 91.7%; Pred. No. 9.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
    |||||
Db 362 AAAAAAAAAAGCATGACTGTGACA 365

RESULT 8
AG880409/c      888 bp      DNA      linear      GSS 03-NOV-2004
LOCUS           Oryza sativa (indica cultivar-group) genomic DNA, BAC end sequence,
DEFINITION      BAC clone:K0298H08_R, genomic survey sequence.
ACCESSION       AG880409
VERSION         AG880409.1 GI:55346670
KEYWORDS        GSS.
SOURCE          Oryza sativa (indica cultivar-group)
ORGANISM        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Euhatortideae; Oryzaceae; Oryza.
REFERENCE
  1 Katagiri,S., Wu,J., Ito,Y., Karasawa,M., Shibata,M., Kanamori,H.,
    Katayose,Y., Naniki,N., Matsumoto,T. and Sasaki,T.
    End Sequencing and Chromosomal in silico Mapping of BAC Clones
    Derived from an indica Rice Cultivar, Kasalath
    Breeding Science 54, 273-279 (2004)
  2 (bases 1 to 888)
    Sasaki,T., Matsumoto,T. and Wu,J.
    Direct Submission
    Submitted (29-OCT-2004) Takuji Sasaki, National Institute of
    Agrobiological Sciences, Rice Genome Research Program, Kannondai
    2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
    (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rsgp.dna.affrc.go.jp/,
    Tel:81-298-38-7441, Fax:81-298-38-7468)
    The orientation of the sequence is from Spe side of the BAC clone.

COMMENT
FEATURES
  Source
    1..888
      /organism="Oryza sativa (indica cultivar-group)"
      /mol_type="genomic DNA"
      /cultivar="Kasalath"
      /db_xref="taxon:39946"
      /clone="K0298H08_R"

ORIGIN
Query Match      83.2%; Score 20.8; DB 10; Length 888;
Best Local Similarity 91.7%; Pred. No. 9.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
    |||||
Db 86 AAAAAAAAAAGCATGACTGTGACA 63

RESULT 9
BU465605/c      912 bp      mRNA      linear      EST 29-NOV-2002
LOCUS           BU465605
DEFINITION      603776458F1 CSEORBN19 Gallus gallus cDNA clone CHEST718c15 5', mRNA
sequence.
ACCESSION       BU465605
VERSION         BU465605.1 GI:25955079
KEYWORDS        EST.

```

```

SOURCE
ORGANISM        Gallus gallus (chicken)
                Gallus gallus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
                Phasianinae; Gallus.
REFERENCE
  1 (bases 1 to 912)
    Beartman,P.E., Sanz-Exquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
    Feng,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
    A Comprehensive Collection of Chicken CDNAS
    Curr. Biol. 12 (22), 1965-1969 (2002)
    12445392
    Contact: Simon Hubbard
    Department of Biomolecular Sciences
    University of Manchester Institute of Science and Technology
    (UMIST)
    PO Box 88, Manchester, M60 1QD, UK
    Tel: 01612008930
    Fax: 01612360409
    Email: Simon.Hubbard@umist.ac.uk.

FEATURES
  Source
    1..912
      /organism="Gallus gallus"
      /mol_type="mRNA"
      /strain="Laver"
      /db_xref="taxon:9031"
      /clone="CHEST718c15"
      /sex="Female"
      /dev_stage="adult"
      /lab_host="DH10B"
      /clone_lib="CSEORBN19"
      /note="Organ: ovary; Vector: pbluescript II KS(+); Site_1:
        ECORI; Site_2: NotI; This normalized library was
        constructed from 1 million independent clones. cDNA
        synthesis was initiated using an oligo(dT) primer, using
        methylated C in the first strand synthesis reaction.
        Following this first strand reaction, double-stranded cDNA
        was blunt-ended, ligated to NotI adapters, digested with
        EcoRI, size-selected, and cloned into the NotI and EcoRI
        compatible sites of a custom modified MCS of the
        pbluescript (KS+) vector. The library was normalized in 2
        rounds using conditions adapted from Soares et al., PNAS
        (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
        (1996): 791, except that a significantly longer
        reannealing hybridization was used."

ORIGIN
Query Match      83.2%; Score 20.8; DB 5; Length 912;
Best Local Similarity 91.7%; Pred. No. 9.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACA 24
    |||||
Db 732 AAAAAAAAAAGCATGACTGTGANA 709

RESULT 10
W11074/c      426 bp      mRNA      linear      EST 02-OCT-1997
LOCUS           W11074
DEFINITION      ma76609.r1 Soares mouse p3JNMF19.5 Mus musculus cDNA clone
IMAGE:316648 5', mRNA sequence.
ACCESSION       W11074
VERSION         W11074.1 GI:1285379
KEYWORDS        EST.
SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 426)
    Maira,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
    Geisel,S., Kucaba,T., Lacy,M., Le M., Martin,J., Morris,M.,
    Scheinberg,K., Stepcie,M., Tan,F., Underwood,K., Moore,B.,
    Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

```

TITLE The Maenu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maena M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through JLNLI; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI: 207264

Seq primer: ETPprimer
High quality sequence stop: 425.
Location/Qualifiers
1. .426
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:316648"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Sceres mouse p3NMF19.5"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAGTGGAGCGCGCATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Sceres and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

FEATURES

source

ORIGIN

Query Match 81.6%; Score 20.4; DB 8; Length 426;
Best Local Similarity 95.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
|||||
DB 60 AAAAAAAAAAGCATGACTGTGA 39

RESULT 11
DE112709/c 434 bp DNA linear GSS 02-AUG-2005
LOCUS DE112709
DEFINITION Oryzias latipes DNA, reverse end of BAC clone: Md0156F21, genomic
survey sequence.
ACCESSION DE112709
VERSION DE112709.1 GI:71632710
KEYWORDS GSS.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.

REFERENCE 1 Shimizu, N., Asakawa, S., Shimizu, A. and Sasaki, T.
The BAC end sequence of Oryzias latipes
JOURNAL Published Only in Database (2005)
REFERENCE 2 (bases 1 to 434)
AUTHORS Shimizu, N., Asakawa, S., Sasaki, T. and Shimizu, A.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-2005) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku,
Tokyo, 160-8582, Japan (E-mail: nshimizu@mb.med.keio.ac.jp,
Tel: 81-3-3351-2370, Fax: 81-3-3351-2370)

FEATURES

source

1. .434

/organism="Oryzias latipes"
/mol_type="genomic DNA"
/db_xref="taxon:8090"
/clone="Md0156F21"
/clone_lib="Medaka Hdr BAC library"
/note="This sequence is reverse end of BAC clone
Md0156F21."

ORIGIN

Query Match 81.6%; Score 20.4; DB 11; Length 434;
Best Local Similarity 95.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
|||||
DB 331 AAAAAAAAAAGCATGACTGTGA 310

RESULT 12
BG588669 436 bp mRNA linear EST 12-APR-2001
LOCUS BG588669
DEFINITION EST490478 MHRP- Medicago truncatula cDNA clone pMHRP-57M7, mRNA
sequence.
ACCESSION BG588669
VERSION BG588669.1 GI:13606809
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicotyledons;
rosids; eustroide I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.

REFERENCE 1 (bases 1 to 436)
Harrison, M.J., Liu, J., Town, C.D., Van Aken, S., Uteback, T., Cho, J.
and Fraser, C.M.
ESTs from phosphate-starved roots of Medicago truncatula, 2001
JOURNAL Unpublished (2001)
CONTACT: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
The Samuel Roberts Noble Foundation: N387420e TIGR sequence name:
MTHC767K More information is available at: <http://www.medicago.org>
Seq primer: SKmod (CTA GAA CTA gtc gat cc).
Location/Qualifiers

1.436
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pMHRP-57M7"
/tissue_type="roots"
/dev_stage="phosphate-starved"
/clone_lib="MHRP-"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; At the trifoliolate stage, M. truncatula plants were
transplanted to phosphate-free sand and grown for a
further 30 days. During this period, they were fertilized
twice weekly with 1/2 Hoaglands solutions containing 20mM
potassium phosphate. cDNA was prepared from polyA+
enriched RNA. The cDNA was directionally ligated into the
Unizap XR vector from Stratagene and packaged using
Gigapack III Gold packaging extracts. Plasmids containing
cDNA inserts were excised from the recombinant lambda-zap
phage using Ex-assist helper phage and propagated in
XL0R cells."

FEATURES
source
1.436
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pMHRP-57M7"
/tissue_type="roots"
/dev_stage="phosphate-starved"
/clone_lib="MHRP-"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; At the trifoliolate stage, M. truncatula plants were
transplanted to phosphate-free sand and grown for a
further 30 days. During this period, they were fertilized
twice weekly with 1/2 Hoaglands solutions containing 20mM
potassium phosphate. cDNA was prepared from polyA+
enriched RNA. The cDNA was directionally ligated into the
Unizap XR vector from Stratagene and packaged using
Gigapack III Gold packaging extracts. Plasmids containing
cDNA inserts were excised from the recombinant lambda-zap
phage using Ex-assist helper phage and propagated in
XL0R cells."

ORIGIN

Query Match

81.6%; Score 20.4; DB 2; Length 436;

Best Local Similarity 95.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
|||||
242 AAAAAAAAAAGCATGACTGTGA 263

RESULT 13
BX516673/c 525 bp mRNA linear EST 27-JUN-2003

LOCUS BX516673 Soares mouse p3JNMF19.5 Mus musculus cDNA clone
IMAGE522181; IMAGE:316366, mRNA sequence.

ACCESSION BX516673
VERSION BX516673.1 GI:32299050
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 525)
Heil, O., Ebert, L., Neubert, P., Peters, M., Radelof, U., Schneider, D.
and Korn, B.

TITLE Mouse Unigeneset - RZPD2
JOURNAL Unpublished (2003)
COMMENT Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE522181.

RZPDLIB: I.M.A.G.E. cDNA Clone Collection;
Mouse Unigeneset - RZPD2 (RZPDLIB No.981)
http://www.rzpd.de/CloneCards/cgi-
bin/showlib.pl.cgi?response2libNo=981 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de

FEATURES
source
1..525
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE522181; IMAGE:316366"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3JNMF19.5"
/note="Vector: pTZ19D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(T) primer [5'
TCTTACCACTGTAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pTZ19D vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonalao. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 81.6%; Score 20.4; DB 5; Length 525;
Best Local Similarity 95.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
|||||
44 AAAAAAAAAAGCATGACTGTGA 23

RESULT 14
CB840280 612 bp mRNA linear EST 01-SEP-2003
LOCUS CB840280/c
DEFINITION M15E-0530 MOUSE EMBRYONIC DAY 15.5 EYE Mus musculus cDNA 5', mRNA
sequence.

ACCESSION CB840280
VERSION CB840280.1 GI:34371668
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 612)
Yu, J., Farjo, R., MacNee, S.P., Baehr, W., Stambolian, D.E. and
Swaroop, A.

TITLE Annotation and analysis of 10,000 expressed sequence tags from
developing mouse eye and adult retina
JOURNAL Genome Biol. 4 (10), R65 (2003)
PUBMED 14519200

COMMENT Contact: Swaroop, A.
Department of Ophthalmology and Visual Sciences
Kellogg Eye Center, University of Michigan
540 KEC, 1000 Wall St., Ann Arbor, MI 48105, USA
Tel: 734 615 2246
Fax: 734 647 0228
Email: swaroop@umich.edu.

FEATURES
source
1..612
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/cissue_type="eye"
/clone_lib="MOUSE EMBRYONIC DAY 15.5 EYE"
/note="Vector: pSPORT1"

ORIGIN
Query Match 81.6%; Score 20.4; DB 6; Length 612;
Best Local Similarity 95.5%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
|||||
531 AAAAAAAAAAGCATGACTGTGA 510

RESULT 15
BP766673 637 bp mRNA linear EST 02-DEC-2004

LOCUS BP766673
DEFINITION BP766673 mouse (C57BL/6) pancreatic islet library with
recombination-based method Mus musculus cDNA clone mid13085 3',
mRNA sequence.

ACCESSION BP766673
VERSION BP766673.1 GI:50225371
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 637)
Nishimura, M., Yokoi, N., Miki, T., Horikawa, Y., Yoshioka, H.,
Takeda, J., Ohara, O. and Seino, S.

TITLE Construction of a multi-functional cDNA library specific for mouse
pancreatic islets and its application to microarray
JOURNAL DNA Res. 11 (5), 315-323 (2004)
PUBMED 15747579

COMMENT Contact: Susumu Seino
Division of Cellular and Molecular Medicine
Kobe University Graduate School of Medicine
7-5-1, Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan
Tel: 81-78-382-5360
Fax: 81-78-382-5370

Email: seino@med.kobe-u.ac.jp

FEATURES

Source

```

Location/Qualifiers
1. .637
   /organism="Mus musculus"
   /mol_type="mRNA"
   /strain="C57BL/6"
   /db_xref="taxon:10090"
   /clone="mid13085"
   /sex="male"
   /tissue_type="pancreatic islet"
   /dev_stage="adult"
   /clone_lib="mouse (C57BL/6) pancreatic islet library with
recombination-based method"

```

ORIGIN

Query Match	81.6%	Score 20.4;	DB 3;	Length 637;
Best Local Similarity	95.5%	Pred. No. 1.3e+03;		
Matches 21; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
Db 589 AAAAAAAAAAGCCTGACTGTGA 610

RESULT 16

LOCUS	CL141446	756 bp	DNA	linear	GSS 05-JUN-2004
DEFINITION	ISB1_118011_T7.1 ISB1 <i>Xenopus tropicalis</i> genomic clone ISB1-118011, genomic survey sequence.				

ACCESSION	CU141446
VERSION	CU141446.1
KEYWORDS	GI:40635081
SOURCE	GSS.
ORGANISM	Xenopus tropicalis (western clawed frog)
	Xenopus tropicalis

REFERENCE

AUTHORS Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.

TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)

COMMENT
Contact: Richard K Wilson

Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGGG

Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 623.

FEATURES

Source

```
1..756
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-118011"
/clone_1b="ISB1"
/notes="Vector: pBeloBac11; ISB-1 Xenopus tropicalis BAC; library Segment 1"
```

ORIGIN

Query Match	81.6%	Score 20.4	DB 10	Length 756
Best Local Similarity	95.5%	Pred. No. 1.4e+03		
Matches 21; Conservative	0	Mismatches 1	Indels 0	Gaps 0

QY	1	AAAAAAAAAGCATGACTGTGA	22
Db	258	AAAAAAAAAGCAAGACTGTGA	279

RESULT 17

CA920541/C

LOCUS	CA920541	816 bp	mRNA	linear	EST 09-MAY-2003
DEFINITION	EST658259	MTUS	Medicago truncatula	CDNA clone MTUS-29E3,	mRNA
				sequence.	

ACCESSION	CA920541
VERSION	CA920541.1
	GI:27407471

KEYWORDS	EST.
SOURCE	Medicago truncatula (barrel medic)

ORGANISM

REFERENCE
1 (bases 1 to 816)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; eudicotyledons: core eudicotyledons
rosids; eurosid I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 816)

AUTHORS

UTTERBACK, T., CHEUNG, F. and FRASER, C.M.:
The Medicago truncatula 6K unigene set: cDNA clones selected and
re-arrayed from various libraries
Unpublished (2002)

COMMENT

Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave., St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu

FEATURES

Sources

```

/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="MTUS-2983"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_lib="MTUS"
/notes="Vector: pBluescript SK-, Site_1: EcoRI, Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the UniZap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-6981st
helper phage and propagated in XLOLR cells."

```

ORIGIN

Query Match	81.6%	Score 20.4;	DB 6;	Length 816;
Best Local Similarity	95.5%;	Pred. No. 1.4e+03;		
Matches 21; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY 1 AAAAAAAAAAGCATGACTGTGA 22
| | | | |
Db 186 AAAAAACACAGCATGACTGTGA 165

RESULT 18

LOCUS	CK787550	821 bp	mRNA	linear	EST 25-FEB-2004
DEFINITION	AGENCOURT 18671903 NIH MGC 185 Mus musculus cDNA clone IMAGE:30844122 5', mRNA sequence.				

ACCESSION	CK787550
VERSION	CK787550.1
	GI:42799545

KEYWORDS

SOURCE

ORGANISM

REFERENCE
AUTHORS
Nih-MGC <http://mgc.ncl.nih.gov/>.
1 (bases 1 to 821)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Daniela S. Gerhard, Ph.D.

FEATURES	Location/Qualifiers
source	1. .821

ORIGIN

Site 2: Not1; RNA obtained from 5 normal wild-type mice cDNA was primed using oligo-dT primer: 5'-pGACTTATTGATGCGAGCGCGCCCT(1)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection 1.4 kb resulted in an average insert size of 1.2 kb. This primary, nanomquantity library is normalized to CoE5 (non-normalized primary library is NIH_MGC_230) and was constructed by Express Genomics (Frederick, MD). Note: this is a NIH_MGC library"

Db 579 AAAAAAAAAAGCCTGACTGTGA 600

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 892)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

```

Plate: L1AM1643 row: 1 column: 2
High quality sequence start: 41
High quality sequence stop: 865.
Location/Qualifiers
1. .892

```

ORIGIN

Query Match	81.6%;	Score 20.4;	DB 3;	length 892;
Best Local Similarity	95.5%;	Pred. No. 1.4e+03;		
Matches 21; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

QY	1	AAAAAAAAA	GCATGCTGA	22
Db	562	AAAAAAAAA <td>AGCTGACTGTA</td> <td>583</td>	AGCTGACTGTA	583
RESULT 20	AA936662			
LOCUS	AA936662			
DEFINITION	AA936662	210 bp	mrna	linear
ACCESSION	U187002.1	NCI CGAP Kids Homo sapiens	CDNA clone IMAGE:1536579	3'
VERSION	AA936662	similar to gb:M75126	HEXOKINASE, TYPE I (HUMAN) ;	mrna sequence.
KEYWORDS	AA936662.1	GI:3094696		
SOURCE	EST.			
ORGANISM	Homo sapiens			
	Homo sapiens (human)			
	Homo sapiens			

FEATURES

```

Trace considered overall poor quality
Insert Length: 586      Std Error: 0.00
Seq primer: -40m3 fwd. BT from Amersham
High quality sequence stop: 1.
Location/Qualifiers
    1. . 210
       /organism="Homo sapiens"
       /mol_type="mRNA"
       /db_xref="taxon:9606"
       /clone="IMAGE:1536579"

```

```

/issue type="2 pooled tumors (clear cell type)"
/lab host="DH10B"
/clone lib="NCI CGAP Kids"
/notes="Organ: Kidney; Vector: pVT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGAGAGATTCGGCGCGCAATATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pVT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldi. "
```

ORIGIN

Query Match 80.8%; Score 20.2; DB 1; Length 210;
Best Local Similarity 88.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
140 AAAAAAAAAAGCAAGCATGTTACAC 164

RESULT 21
LOCUS A0097896 330 bp DNA linear GSS 27-AUG-1998
DEFINITION HS_3036_B2_B08_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3036 Col=16 Row=D, genomic survey
sequence.
ACCESSION A0097896
VERSION A0097896.1 GI:3468925
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 330)
Keller, A., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10449764
Contact: Mahalax G., Wallace J.C., Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3036 row: D column: 16
Class: BAC ends
High quality sequence stop: 330.
Location/Qualifiers
1. .330
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="Plate=3036 Col=16 Row=D"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: Sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 330;
Best Local Similarity 88.0%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25

```

Db 136 AAAAAAAAAAATGACTGTACAC 112
```

RESULT 22
LOCUS CW730653 404 bp DNA linear GSS 05-NOV-2004
DEFINITION MARC_777462 CHORI-240-437J9 Bos taurus genomic clone 1B5, genomic
survey sequence.
ACCESSION CW730653
VERSION CW730653.1 GI:55439222
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Bovidae; Bovinae; Bos.
1 (bases 1 to 404)
Smith, T.P.L., Stone, R.T., Keeler, J.W., Snelling, W.M. and Harhay, G.P.
SNP discovery in cattle based on low coverage sequencing of BAC
clones
Unpublished (2004)
Other GSSes: MARC_777558
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smtlh@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Seq primer: AATTAACTCTCACTAAAGGC
Class: BAC subclone.
Location/Qualifiers
1. .404
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone_lib="1B5"
/sex="Male"
/tissue_type="Blood"
/clone_lib="CHORI-240-437J9"
/notes="Vector: pBLUESCRIPT SK-; Site 1: BamHI, Site 2:
BamHI; BAC DNA was digested with Sau3A. Fragments were
sized to 800-1200 bases and subcloned into pBLUESCRIPT
SK-."

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 404;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
141 AAAAAAAAAAGCATGACGCAAC 165

RESULT 23
LOCUS AA931550 412 bp mRNA linear EST 06-MAY-1998
DEFINITION O056B08.81 NCI_CGAP_Lus Homo sapiens cDNA clone IMAGE:1570167 3',
mRNA sequence.
ACCESSION AA931550
VERSION AA931550.1 GI:3085936
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 412)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmer-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www-bio.lnl.gov/hbbrp/image/image.html
Insert Length: 739 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 402.

FEATURES

source

1. .412
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1570167"
/tissue_type="carcinoid"
/lab_host="DH10B"
/clone_lib="NCI-CGAP LUS"
/note="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 80.8%; Score 20.2; DB 1; Length 412;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
121 AAAAAAAAAAGCATGACTGTGAAACAC 145

Db 121 AAAAAAAAAAGCATGACTGTGAAACAC 145

RESULT 24 CM730610 446 bp DNA linear GSS 05-NOV-2004
LOCUS CM730610 MARC.777403 CHORI-240-43739 Bos taurus genomic clone 1G1, genomic survey sequence.

DEFINITION CM730610

ACCESSION CM730610

VERSION CM730610.1 GI:55439136

KEYWORDS GSS.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 446)
Smith,T.P.L., Stone,R.T., Keele,J.W., Snelling,W.M. and Harhay,G.P. SNP discovery in cattle based on low coverage sequencing of BAC clones

AUTHORS Smith,T.P.L., Stone,R.T., Keele,J.W., Snelling,W.M. and Harhay,G.P.

TITLE Unpublished (2004)

JOURNAL Other_GSSs: MARC.777499

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smitht@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross match v0.990329.
Seq primer: AATTAAACCTCCTCAAAAGG

Class: BAC subclone.

FEATURES

source

1. .446
Location/Qualifiers
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="1G1"
/sex="Male"
/tissue_type="Blood"
/clone_lib="CHORI-240-43739"
/note="Vector: pBLUESCRIPT SK-; Site 1: BamHI, Site 2: BamHI, BAC DNA was digested with Sau3A. Fragments were sized to 800-1200 bases and subcloned into pBLUESCRIPT SK-."

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 446;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
142 AAAAAAAAAAGCATGACGACGACAC 166

Db 142 AAAAAAAAAAGCATGACGACGACAC 166

RESULT 25 CK059387 447 bp mRNA linear EST 05-FEB-2005
LOCUS CK059387/c 59099rsiek.16383.y1 Oryza sativa cv. PA64s panicle fertile cDNA
DEFINITION library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.

ACCESSION CK059387

VERSION CK059387.1 GI:58670701

KEYWORDS EST.

SOURCE Oryza sativa (indica cultivar-group)

ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 447)
Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Li,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.

AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Li,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.

TITLE The genomes of Oryza sativa: A History of Duplications

JOURNAL PLoS Biol. 3 (2), e38 (2005)

COMMENT 15685292
Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 447
POLYA-No.

FEATURES

source

1. .447
Location/Qualifiers
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="PA64s"
/db_xref="taxon:39946"
/tissue_type="panicle"
/cell_type="fertile"
/dev_stage="heading/flowering"
/clone_lib="Oryza sativa cv. PA64s panicle fertile cDNA library"

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 447;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
 |||||
 169 AAAAAAAAAACATCACTGTGACAC 145

RESULT 26
 CB641312 510 bp RNA linear EST 08-APR-2003
 LOCUS OSJNEA18L09.f OSJNEA Oryza sativa (japonica cultivar-group) cDNA
 DEFINITION
 CB641312
 VERSION CB641312 GI:29636303
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS Jantasuriyarat, C., Gowda, M., Haller, K., Hatfield, J., Lu, G., Stahlberg, E., Zhou, B., Li, H., Kim, H., Yu, Y., Dean, R. A., Wang, R. A., Soderlund, C., and Wang, G. L.
 TITLE Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction
 JOURNAL Plant Physiol. 138 (1), 105-115 (2005)
 PUBMED 15886683
 COMMENT Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: twing@genome.arizona.edu
 PCR PRIMERS
 FORWARD: atc agc ggc cgc gat cc
 BACKWARD: aat taa ccc tca cta aag g
 Plate: 18 row: 1 column: 09
 Seq primer: atc agc ggc cgc gat cc.
 Location/Qualifiers
 1. 510
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEA18L09"
 /tissue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_1lb="OSJNEA"
 /note="Vector: pbluescript II KS +; Site_1: EcoRI; Site_2: XhoI; 6 hrs after inoculation with Rice Blast (Ch 86061)"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 6; Length 510;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
 |||||
 375 AAAAAAAAAACATCACTGTGACAC 351

RESULT 27
 CW730743 525 bp DNA linear GSS 05-NOV-2004
 LOCUS MARC.77569 CHORI-240-437J9 Bos taurus genomic clone 1M5, genomic survey sequence.

ACCESSION CW730743
 VERSION CW730743.1 GI:55439411
 KEYWORDS GSS.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 525)
 Smith, T. P. L., Stone, R. T., Keeler, J. W., Snelling, W. M. and Harhay, G. P.
 TITLE SNP discovery in cattle based on low coverage sequencing of BAC clones
 JOURNAL Unpublished (2004)
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smtlh@mail.marc.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329.
 Seq primer: GTAATACGACTCATCTAGCG
 Class: BAC subclone.
 Location/Qualifiers
 1. 525
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"
 /clone="1M5"
 /sex="Male"
 /tissue_type="blood"
 /clone_1lb="CHORI-240-437J9"
 /note="Vector: pBLUESCRIPT SK-; Site 1: BamHI; Site 2: BamHI; BAC DNA was digested with Sau3A. Fragments were sized to 800-1200 bases and subcloned into pBLUESCRIPT SK-."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 10; Length 525;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
 |||||
 141 AAAAAAAAAACATCACTGTGACAC 165

RESULT 28
 CV844822 556 bp mRNA linear EST 17-NOV-2004
 LOCUS ID0AEB11AF05RM1 ID0AEB Acyrthosiphon pisum cDNA clone ID0AEB11AF05
 DEFINITION 5', mRNA sequence.
 ACCESSION CV844822
 VERSION CV844822.1 GI:55810721
 KEYWORDS EST.
 SOURCE Acyrthosiphon pisum (pea aphid)
 ORGANISM Acyrthosiphon pisum
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosciphini; Acyrthosiphon.
 1 (bases 1 to 556)
 Sabater-Munoz, B., Legault, F., Bonhomme, J., Dang, P., Dosset, C., Duclet, A., Gauthier, J. P., Hunter, W., Martinez-Torres, D., Moysa, A., Nakabachi, A., Prunier-Leterme, N., Rabhe, Y., Shigenobu, S., Simon, J. C., Stern, D., Wincker, P. and Tagu, D.
 TITLE Annotated ESTs of the pea aphid
 JOURNAL Unpublished (2004)
 COMMENT Contact: D. Tagu
 INRA Rennes
 UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France
 Tel: +33.2.23.48.51.65
 Fax: +33.2.23.48.51.50

PCR Primers
 FORWARD: CAGGAACAGCTATGACC
 Plate: 11A Row: F Column: 5.
 Location/Qualifiers

FEATURES

Source

1..556
 /organism="Acyrtosiphon pisum"
 /mol_type="mRNA"
 /cultiivar="yr2"
 /db_xref="taxon:7029"
 /clone="ID0AEE11AF05"
 /tissue_type="antennae"
 /dev_stage="L3"
 /lab_host="XLI-Blue"
 /clone_lib="ID0AEE"
 /note="Vector: pBS-SKminus; Site 1: EcoRI, Site 2: XhoI;
 Sample name: ID0AEE; Plant growth place: INRA Rennes, UMR
 Bio3P, 35327, 35653 le Rheu Cedex France; Soil
 conditions: Soil; Sowing date: 15/04/2004; Harvesting
 date: 15/04/2004; Description: Aphids inoculated on
 one-week old *Vicia faba* under non-sterile conditions. A.
 pisum YR2 is holocyclic, i.e. able to change its
 reproductive mode under short photoperiods (sexual) versus
 long photoperiods (clonal). experimental condition: long
 photoperiod (16-hr light/8-hr dark at 18 degC)"

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 556;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25

Db 475 AAAAAAAAAACATGCTGTGATAC 499

RESULT 29
 BH569468/c 557 bp DNA linear GSS 14-DEC-2001
 LOCUS BCB0337T BCB Brassica oleracea genomic clone BCB033, genomic
 DEFINITION Survey sequence.
 ACCESSION BH569468
 VERSION BH569468.1 GI:17821307
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 557)
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
 Utechtack,T.R., Mortman,J.R., White,O.R. and Town,C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)
 15805490
 Other_GSSs: BCB0337T
 Contact: Chris Town
 TIGR

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 PUBMED
 COMMENT

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..557
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO100DH3"
 /db_xref="taxon:3712"
 /clone="BCB033"
 /clone_lib="BCGB"

FEATURES

Source

1..557
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO100DH3"
 /db_xref="taxon:3712"
 /clone="BCB033"
 /clone_lib="BCGB"

/note="Vector: pBOS1; Site 1: BstXI, 2-3 kb sheared
 genomic DNA inserted into pBOS1 using BstXI linkers"

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 557;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25

Db 506 AAAAAAAAAATCATCTGTGAAC 482

RESULT 30
 CK072813/c 578 bp mRNA linear EST 05-FEB-2005
 LOCUS CK072813
 DEFINITION 67030stsem.2811.v1 Oryza sativa cv. Pa646 panicle fertile cDNA
 library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
 sequence.
 ACCESSION CK072813 GI:58684126
 VERSION CK072813
 KEYWORDS EST.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 578)
 Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
 Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
 Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
 Li,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
 Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
 Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
 Wu,S. and Liu,J.
 The Genomes of Oryza sativa: A History of Duplications
 Plos Biol. 3 (2), e38 (2005)
 15685292
 Contact: Yan Zhou
 Bioinformatics Department
 Hangzhou Genomics Institute
 No.51 Zhijiang Road, Hangzhou 310008, China
 Tel: 86-571-56805886
 Fax: 86-571-56805884
 Email: zhouyan@genomics.org.cn
 Seq primer: M13 Forward
 High quality sequence stop: 578
 POLYA=No.

REFERENCE

AUTHORS

TITLE
 JOURNAL
 PUBMED
 COMMENT

Contact: Yan Zhou
 Bioinformatics Department
 Hangzhou Genomics Institute
 No.51 Zhijiang Road, Hangzhou 310008, China
 Tel: 86-571-56805886
 Fax: 86-571-56805884
 Email: zhouyan@genomics.org.cn
 Seq primer: M13 Forward
 High quality sequence stop: 578
 POLYA=No.

FEATURES

Source

1..578
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultiivar="Pa646"
 /db_xref="taxon:39946"
 /tissue_type="panicle"
 /cell_type="fertile"
 /dev_stage="heading/flowering"
 /clone_lib="Oryza sativa cv. Pa646 panicle fertile cDNA
 library"

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 578;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25

Db 423 AAAAAAAAAACATCTGTGACAC 399

RESULT 31
 CW755031/c 579 bp DNA linear GSS 09-NOV-2004
 LOCUS CW755031

DEFINITION OG_BBa0057K24.r OG_BBa Oryza glaberrima genomic clone OG_BBa0057K24
3', genomic survey sequence.
ACCESSION CW755031
VERSION CW755031.1 GI:55593701
KEYWORDS GSS.
SOURCE Oryza glaberrima (African rice)
ORGANISM Oryza glaberrima
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 579)
AUTHORS Kim H., Yu Y., Wiscocki, M., Byrne, M., Stum, D., Smart, D., Rao, K.,
Luo, M., Jettly, R., Kudrna, D., Muller, C., Hatfield, J., Soderlund, C.,
and Wing, R.
TITLE OMAP
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TTA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0057 row: K column: 24
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
Location/Qualifiers
1..579
/organism="Oryza glaberrima"
/mol_type="genomic DNA"
/db_xref="taxon:4538"
/clone="OG_BBa0057K24"
/issue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OG_BBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 579;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
|||||
560 AAAAAAAAAATCATGCTGTGACAC 536

RESULT 32
BU079563 582 bp mRNA linear EST 29-SEP-2003
LOCUS BU079563 NTBB Mochii normalized Xenopus tailbud library Xenopus
DEFINITION laevis cDNA clone XL072m08 3', mRNA sequence.
ACCESSION BU079563
VERSION BU079563.1 GI:17524479
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 582)
AUTHORS Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
Kohara, Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856

FEATURES
source
1..582
Location/Qualifiers
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL072m08"
/issue_type="whole embryo"
/dev_stage="stage 25"
/clone_lib="NTBB Mochii normalized Xenopus tailbud library"

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 582;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
|||||
163 AAAAAAAAAAGCATGCTTTGACCC 139

RESULT 33
AQ328946 611 bp DNA linear GSS 08-JAN-1999
LOCUS AQ328946
DEFINITION nbx0044G21f CUGI Rice BAC library Oryza sativa (japonica
cultivar-group) genomic clone nbx0044G21f, genomic survey
sequence.
ACCESSION AQ328946
VERSION AQ328946.1 GI:4120796
KEYWORDS GSS.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 611)
AUTHORS Wing, R. A. and Dean, R. A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACCTCACTAATAGCG
Class: BAC ends
High quality sequence strop: 207.
Location/Qualifiers
1..611
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbx0044G21f"
/issue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII. Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small

genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 611;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 264 AAAAAAAAAACATCACTGTGACAC 240

RESULT 34
CL820341/c 653 bp DNA linear GSS 09-AUG-2004
LOCUS OR_C8A0038018.r OR_C8A Oryza rufipogon genomic clone OR_C8A0038018
DEFINITION 3, genomic survey sequence.
ACCESSION CL820341
VERSION CL820341.1 GI:51065951
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Burrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 653)
Kim,H., Yu,Y., Wisnietki,M., Yost,D., Stum,D., Rao,K., Luo,M.,
Jecty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
Wing,R.
OMAP project
Unpublished (2004)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0038 row: O column: 18
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..653
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_C8A0038018"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_C8A"
/note="Vector: pAGT1BAC1; Site 1: HindIII; Site 2: HindIII;
dkf treated 36 hrs before harvest"

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 653;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25

Db 560 AAAAAAAAAATCATGACTGTGACAC 536

RESULT 35
CE223739 664 bp DNA linear GSS 25-SEP-2003
LOCUS tigr-gss-dog-17000326865228 Dog Library Canis familiaris genomic,
DEFINITION 1gfr-gss-dog survey sequence.
ACCESSION CE223739
VERSION CE223739.1 GI:35379454
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 664)
Kirkness,E.F., Batina,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Classes: shotgun.

FEATURES

source

Location/Qualifiers
1..664
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 664;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGACTGTGACAC 25
Db 548 AAAAAAAAAAGATGACTGGAAAC 572

RESULT 36

BE212779 716 bp mRNA linear EST 30-JUN-2000
LOCUS IPBRN00803 Brain cDNA library Ictalurus punctatus cDNA 5', mRNA
DEFINITION sequence.

ACCESSION BE212779
VERSION BE212779.1 GI:8844436
KEYWORDS EST.
ORGANISM Ictalurus punctatus (channel catfish)
Ictalurus punctatus
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Ictaluridae; Ictalurus.

REFERENCE 1 (bases 1 to 716)
Ju,Z., Karsl,A., Kocbas,A., Patterson,A., Li,P., Cao,D., Dunham,R.
and Liu,Z.
Transcriptome analysis of channel catfish: I. genes and expression
profiles from the brain
Unpublished (2000)
Contact: Liu, Z.J.
Fish Molecular Genetics and Biotechnology

TITLE
JOURNAL
COMMENT

Auburn University
203 Swingle Hall, Department of Fisheries, Auburn, AL 36849, USA
Tel: 334 844 4054
Fax: 334 844 9208
Email: zliu@acesag.auburn.edu
Seq primer: M13 Reverse.

FEATURES
source
Location/Qualifiers

1. 716
/organism="Ictalurus punctatus"
/mol_type="mRNA"
/db_xref="taxon:7998"
/clone_lib="Brain cDNA library"
/note="Organ: Brain; Vector: pSport1; Site_1: NCI;
Site_2: Sall"

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 716;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
|||||
Db 594 AAAAAACAACGATGACTGTGACAC 618

RESULT 37
CMB70653 749 bp DNA linear GSS 12-FRB-2005
LOCUS sh2c3-43.9_070.ab1 Whole-genome shotgun library of the elephant
DEFINITION shark (aka elephant fish) Callorhynchus milii genomic, genomic
survey sequence.

ACCESSION CMB70653
VERSION CMB70653.1 GI:59697288
KEYWORDS GSS.
SOURCE Callorhynchus milii (elephantfish)
ORGANISM Callorhynchus milii (elephantfish)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
Holcephali; Chimaeriformes; Callorhynchidae; Callorhynchus.

REFERENCE 1 (bases 1 to 749)
Venkatesh, B., Tay, A., Dandona, N., Patil, J. G. and Brenner, S.
A compact cartilaginous fish model genome
Curr. Biol. 15 (3), R82-R83 (2005)
15694293

COMMENT

Contact: Venkatesh B
Molecular Genetics Lab
Institute of Molecular and Cell Biology
61 Biopolis Drive, Singapore 138673
Tel: 65 6586 9571
Fax: 65 6779 1117
Email: mcbv@imcb.a-star.edu.sg
Whole-genome shotgun sequences of the elephant shark (aka elephant
fish)
Class: shotgun.

FEATURES
source
Location/Qualifiers

1. 749
/organism="Callorhynchus milii"
/mol_type="genomic DNA"
/db_xref="taxon:7868"
/sex="Male"
/tissue_type="Testis"
/clone_lib="Whole-genome shotgun library of the elephant
shark (aka elephant fish)"

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 749;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
|||||
Db 453 AAAAAACAACGATGACTGTGACAC 477

RESULT 38
AQ329214/c 752 bp DNA linear GSS 08-JAN-1999
LOCUS nbxb0044H21f CUGI Rice BAC Library Oryza sativa (japonica
DEFINITION cultivar-group) genomic clone nbxb0044H21f, genomic survey
sequence.

ACCESSION AQ329214
VERSION AQ329214.1 GI:4121064
KEYWORDS GSS.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 752)
Wing, R. A. and Dean, R. A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: TAAATGACTCACTATAGCG
Class: BAC ends

High quality sequence strop: 296.
Location/Qualifiers

FEATURES
source

1. 752
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbxb0044H21f"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBlotSC11; Site 1: HindIII; Site 2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9%. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 752;
Best Local Similarity 88.0%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
|||||
Db 265 AAAAAACAACCATCTCTGACAC 241

RESULT 39
CMB71063/c 760 bp mRNA linear EST 09-APR-2003
LOCUS OSJNB04H07.f OSJNB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION

ACCESSION Clone OSJNE04H07 5', mRNA sequence.
 VERSION CB671063
 KEYWORDS CB671063.1 GI:29674788
 SOURCE EST.
 ORGANISM *Oryza sativa* (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euphorbiaceae; Oryzae; *Oryza*.
 1 (bases 1 to 760)
 Jantarayarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
 Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Ming,R.A.,
 Soderlund,C. and Wang,G.L.
 Large-scale identification of expressed sequence tags involved in
 rice and rice blast fungus interaction
 Plant Physiol. 138 (1), 105-115 (2005)
 1588683
 COMMENT Contact: Rod Ming
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: rwing@genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 04 row: H column: 07
 Seq primer: gta aaa cga cgg cca gtc.
 FEATURES
 source
 1..760
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNE04H07"
 /tissue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_id="OSJNE"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (70-15)"
 ORIGIN
 Query Match 80.8%; Score 20.2; DB 6; Length 760;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
 |||||
 Db 531 AAAAAAAAAACATCACTGTGACAC 507
 |||||
 RESULT 40
 LOCUS BH125511 791 bp DNA linear GSS 19-JUN-2001
 DEFINITION RPCI-24-289J21.TV RPCI-24 Mus musculus genomic clone
 BH125511
 RPCI-24-289J21, genomic survey sequence.
 VERSION BH125511
 KEYWORDS BH125511.1 GI:14969023
 GSS.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 791)
 Zhao,S., Nierman,W., Malek,J., Shatman,S., Akincet,B., Levins,M.,
 Teegaye,G., Geer,K., Kroll,M., Shvartsbeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)

COMMENT Other GSSs: RPCI-24-289J21.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 289 row: J column: 21
 Seq primer: T7
 Class: BAC ends.
 FEATURES
 source
 Location/Qualifiers
 1..791
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-289J21"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /clone_id="RPCI-24"
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 ORIGIN
 Query Match 80.8%; Score 20.2; DB 9; Length 791;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGACAC 25
 |||||
 Db 242 AAAAAATAAAGCTTGTCTGTGACAC 266
 |||||
 RESULT 41
 LOCUS BX163609 796 bp DNA linear GSS 13-MAR-2003
 DEFINITION Danio rerio genomic clone DKEY-12414, genomic survey sequence.
 ACCESSION BX163609
 VERSION BX163609.1 GI:27995128
 KEYWORDS GSS.
 SOURCE Danio rerio (zebrafish)
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 796)
 Humphray,S.J., Huckle,E. and Durham,J.L.
 Direct Submision
 Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Unpublished
 This sequence was generated from the T7 end of BAC 12414. 12414 is
 part of the Daniokey BAC library created by R. Plasterk and N.V.
 Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.
 Location/Qualifiers
 1..796
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEY-12414"
 /tissue_type="testis"
 /note="vector pindigoBAC-536"
 ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 796;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGCTGTGACAC 25
 |||
 581 AAAAAAAAAAGCATGCTGTGACAC 605

RESULT 42
 LOCUS BX156386 808 bp DNA linear GSS 28-JAN-2003
 DEFINITION Danio rerio genomic clone DKRY-15IG11, genomic survey sequence.
 ACCESSION BX156386
 VERSION BX156386.1 GI:27987973
 KEYWORDS GSS.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 808)
 Humphrey,S.J., Huckle,E. and Durham,J.L.
 Direct Submission (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Unpublished
 This sequence was generated from the SP6 end of BAC 15IG11. 15IG11 is part of the Daniokey BAC library created by R. Plasterk and N.V. Keygene. Further details:
 http://www.sanger.ac.uk/Projects/D_rerio/
 Location/Qualifiers
 source 1..808
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKRY-15IG11"
 /issue_type="Testis"
 /note="vector pindigobAC-536"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 10; Length 808;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGCTGTGACAC 25
 |||
 137 AAAAAAAAAAGCATGCTGTGACAC 161

RESULT 43
 LOCUS CL762965/c 854 bp DNA linear GSS 27-JUL-2004
 DEFINITION OR_BBA0133B19.r OR_BBA Oryza niwara genomic clone OR_BBA0133B19 3', genomic survey sequence.
 ACCESSION CL762965
 VERSION CL762965.1 GI:50720912
 KEYWORDS GSS.
 SOURCE Oryza niwara
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 854)
 Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jerry,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
 OMAP Project
 Unpublished (2004)
 Contact: Rod A. Wing
 Arizona Genomics Institute
 University of Arizona
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595
 Fax: 520 621 1259
 Email: twing@genome.arizona.edu
 PCR Primers
 FORWARD: TAA TAC GAC TCA CTA TAG GG
 BACKWARD: CAC TCA TTA GGC ACC CCA
 Insert Length: 161 Std Error: 0.00
 Plate: 0133 row: E column: 19
 Seq primer: CAC TCA TTA GGC ACC CCA
 Class: BAC ends.

FEATURES
 source 1..854
 Location/Qualifiers
 1..854
 /organism="Oryza niwara"
 /mol_type="genomic DNA"
 /db_xref="taxon:4536"
 /clone="OR_BBA0133B19"
 /issue_type="young leaves"
 /lab_host="DH10B-T1 phage resistant"
 /clone_1lb="OR_BBA"
 /note="vector: pAG1BAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 10; Length 854;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGCTGTGACAC 25
 |||
 92 AAAAAAAAAATCATGCTGTGACAC 68

RESULT 44
 LOCUS DR922466 917 bp mRNA linear EST 02-AUG-2005
 DEFINITION EST1114005 Aquilegia cDNA library Aquilegia formosa x Aquilegia pubescens cDNA clone COLMQ19, mRNA sequence.
 ACCESSION DR922466
 VERSION DR922466.1 GI:71691829
 KEYWORDS EST.
 SOURCE Aquilegia formosa x Aquilegia pubescens
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales; Ranunculaceae; Aquilegia.
 1 (bases 1 to 917)
 Hodges,S.A., Rensink,W., Buell,C.R., Borevitz,J., Kramer,E., Nordborg,M. and Tomkins,J.
 Generation of ESTs from Aquilegia
 Unpublished (2005)
 Other_ESTs: EST1114006
 Contact: Scott Hodges
 Department of Ecology, Evolution and Marine Biology
 University of California, Santa Barbara
 Santa Barbara, CA 93106, USA
 Tel: 805 893 7813
 Fax: 805 893 4724
 Email: hodges@lifesci.ucsb.edu
 Seq primer: TTTTGTTTTGTTTTGTTTT (where N = A, G & C).
 Location/Qualifiers
 1..917
 /organism="Aquilegia formosa x Aquilegia pubescens"
 /mol_type="mRNA"
 /db_xref="taxon:338618"
 /clone="COLMQ19"
 /issue_type="mixed shoot and floral apical meristems, flower buds, leaves and roots"
 /lab_host="DH10B T1 (T1 and T5 phage resistance)"
 /clone_1lb="Aquilegia cDNA library"
 /note="vector: pCMV SPORT6.1; Site_1: EcoRI; Site_2: NotI; F2, F3, and F4 lines of Aquilegia formosa x A. pubescens were grown from seed in greenhouses at UC Santa Barbara. From these plants three sets of tissue were collected: 1) Small flower buds (<10 mm) and very young inflorescences

REFERENCE 1 (bases 1 to 351)
AUTHORS Kirkness, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: shotgun.
FEATURES
source Location/Qualifiers
1..351
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN
Query Match 80.0%; Score 20; DB 10; Length 351;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGACTGT 20
|||||
303 AAAAAAAAAAGCATGACTGT 284

RESULT 48
LOCUS CNS021YAM 886 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ori end of clone 180A22 of library G from Tetradon nigroviridis, genomic survey sequence.
ACCESSION AL219415
VERSION AL219415.1 GI:7878234
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Butelosteii; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetradon.
REFERENCE 1
AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizes, C., Wincker, P., Broccher, P., Quetier, F., Saurin, W. and Weissenbach, J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetradon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Ozoar-Costaz, C., Fizes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
PUBMED 10899143
REFERENCE 3 (bases 1 to 886)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 911 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large

FEATURES
source Location/Qualifiers
1..886
/organism="Tetradon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone_lib="180A22"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG180BA11SP1 end : PUC-ori"

ORIGIN
Query Match 80.0%; Score 20; DB 10; Length 886;
Best Local Similarity 87.5%; Pred. No. 2e+03;
Matches 21; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGACTGTACA 24
|||||
144 AAAAAAAAAAGCATGACTGTACA 167

RESULT 49
LOCUS AZ576423 344 bp DNA linear GSS 06-DEC-2000
DEFINITION A5T-2T00934 Genetrap T47D Human Breast Carcinoma Library Homo sapiens genomic 5', genomic survey sequence.
ACCESSION AZ576423
VERSION AZ576423.1 GI:11562734
KEYWORDS GSS.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Carchinhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 344)
AUTHORS Henkel, G., Liyanage, M., Pralt, E., Huang, D., Riley, M., Bernadino, A., Durick, K. and Pollok, B.
TITLE Exon-trap tags from a T47D Genomescreen (TM) Library
JOURNAL Unpublished (2000)
COMMENT Contact: Greg Henkel
Gene Expression
Auroa Biosciences Corp.
11010 Torreyana Road, San Diego, CA 92121, USA
Tel: 8584048436
Fax: 8584046719
Email: henkel@aurorabio.com
Pools of cells were isolated from a Genomescreen (TM) library. The library of cells was generated by retroviral integration of a gene tagging element consisting of: 1) A promoterless beta-lactamase preceded by a splice acceptor as a reporter for gene expression; 2) A promoter driving neomycin resistance followed by a splice donor to trap downstream exons. 3' RACE from neomycin gene was performed using total RNA from isolated pools. Output was shotgun cloned in pAMP-1 and used to transform DH5-alpha competent bacteria. 5' ends of reported sequences were immediately preceded by splice donor from the trapping construct.
Class: exon-trapped.
FEATURES
source Location/Qualifiers
1..344
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/tissue_type="Carcinoma"
/cell_type="Epithelial"
/cell_line="T47D"
/clone_lib="Genetrap T47D Human Breast Carcinoma Library"
/note="Organ: Breast; Vector: pAMP-1; 3' RACE of total RNA from genetrap pools; shotgun clone in pAMP-1 and used to transform DH5-alpha competent bacteria."

ORIGIN

Query Match 79.2%; Score 19.8; DB 9; Length 344;
 Best Local Similarity 91.3%; Pred. No. 2.2e+03;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGCATGCTGTGACAC 25
 |||||
 DB 250 AAAAATAACATGACTGTGACAC 272

RESULT 50

BH189388/c

LOCUS BH189388 400 bp DNA linear GSS 19-OCT-2001
 DEFINITION 037_P_16-21 SmbAC1 Schistosoma mansoni genomic clone 037P16 5',
 genomic survey sequence.

ACCESSION

BH189388

VERSION

BH189388.1 GI:16301386

KEYWORDS

GSS

SOURCE

Schistosoma mansoni

ORGANISM

Schistosoma mansoni

REFERENCE

Strigoida; Schistosomatidae; Schistosomatidae; Schistosoma.

AUTHORS

Le Paslier, M.C., Pierce, R.J., Merlin, F., Hirai, H., Wu, W.,
 Williams, D.L., Johnston, D., Loverde, P.T., and Le Paslier, D.

TITLE

Construction and characterization of a Schistosoma mansoni
 bacterial artificial chromosome library

JOURNAL

Genomics 65 (2), 87-94 (2000)

PubMed

10783255

COMMENT

Other_GSSs: 037_P_16-rev
 Contact: Pierce RJ

INSERM U 167

Institut Pasteur de Lille

1 rue du Professeur A. Calmette, 59019-Lille, France

Tel: (33) (0)3 20877883

Fax: (33) (0)3 20877888

Email: Raymond.Pierce@pasteur-lille.fr

CNS sequencing ID=DG0AA037DH08CP1

Plate: 037 row: P column: 16

Seq primer: M13 -21 primer

Class: BAC ends

High quality sequence stop: 400.

Location/Qualifiers

FEATURES

1..400

source

/organism="Schistosoma mansoni"

/mol_type="genomic DNA"

/strain="Puerto-Rican"

/db_xref="taxon:6183"

/clone="037P16"

/sex="mixed"

/dev_stage="cercariae"

/lab_host="Biomphalaria glabrata"

/clone_lib="SmbAC1"

/note="Vector: pBelOBAC 11; site 1: Hind III; Partially
 Hind III digested and size-selected S. mansoni cercarial
 DNA was ligated into Hind III digested pBelOBAC 11 vector
 and used to transform E. coli DH10B. The complete library
 contains 23808 clones from 4 independent
 sizing-ligation-transformations. Average insert size
 ranges from 70-127 kb and genome coverage is 7.9-fold."

ORIGIN

Query Match

79.2%; Score 19.8; DB 9; Length 400;

Best Local Similarity

91.3%; Pred. No. 2.3e+03;

Matches

21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

3 AAAAAAAAAAGCATGCTGTGACAC 25

DB

150 AAAAACAAGCATGCTGTGAAC 128

Search completed: December 14, 2005, 07:35:05

Job time: 1762.1 secs

THIS PAGE BLANK (USPTO)

C 97 17 68.0 16963 6 US-10-995-561-13467 Sequence 13467, A
C 98 17 68.0 40000 6 US-10-995-561-13513 Sequence 13513, A
C 99 17 68.0 45038 6 US-10-995-561-13311 Sequence 13311, A
C 100 17 68.0 88116 6 US-10-995-561-13351 Sequence 13351, A
C 101 17 68.0 105550 6 US-10-995-561-13335 Sequence 13325, A
C 102 17 68.0 121001 7 US-11-117-187-190 Sequence 190, App
C 103 17 68.0 138821 7 US-11-121-086-80 Sequence 80, App
C 104 17 68.0 146556 7 US-11-121-086-68 Sequence 68, App
C 105 17 68.0 147700 6 US-10-857-780-3 Sequence 3, App
C 106 17 68.0 150481 7 US-11-112-908-37 Sequence 37, App
C 107 17 68.0 154452 7 US-11-121-086-74 Sequence 74, App
C 108 17 68.0 158692 7 US-11-121-086-30 Sequence 30, App
C 109 17 68.0 166111 7 US-11-112-908-47 Sequence 47, App
C 110 17 68.0 169495 7 US-11-121-086-61 Sequence 61, App
C 111 17 68.0 170837 7 US-11-121-086-97 Sequence 97, App
C 112 17 68.0 170995 7 US-11-121-086-35 Sequence 35, App
C 113 17 68.0 179892 7 US-11-112-908-39 Sequence 39, App
C 114 17 68.0 186442 7 US-11-121-086-104 Sequence 104, App
C 115 17 68.0 203467 7 US-11-121-086-50 Sequence 50, App
C 116 17 68.0 222094 6 US-10-995-561-13244 Sequence 13244, A
C 117 17 68.0 222094 6 US-10-995-561-13244 Sequence 13244, A
C 118 17 68.0 305312 7 US-10-995-561-13236 Sequence 13236, A
C 119 17 68.0 1082144 7 US-11-117-187-211 Sequence 211, App
C 120 17 68.0 1125000 6 US-10-995-561-13286 Sequence 13286, A
C 121 16.8 67.2 25 7 US-11-121-849-76124 Sequence 76124, A
C 122 16.8 67.2 25 7 US-10-995-561-26086 Sequence 26086, A
C 123 16.8 67.2 201 6 US-10-995-561-52126 Sequence 52126, A
C 124 16.8 67.2 201 6 US-10-995-561-55546 Sequence 55546, A
C 125 16.8 67.2 201 6 US-10-995-561-70296 Sequence 70296, A
C 126 16.8 67.2 3553 6 US-10-750-185-40225 Sequence 40225, A
C 127 16.8 67.2 27403 6 US-10-995-561-13419 Sequence 13419, A
C 128 16.8 67.2 45038 6 US-10-995-561-13311 Sequence 13311, A
C 129 16.8 67.2 165627 7 US-11-121-086-89 Sequence 89, App
C 130 16.8 67.2 179892 7 US-11-112-908-39 Sequence 39, App
C 131 16.8 67.2 24496 6 US-10-995-561-13327 Sequence 13327, A
C 132 16.8 66.4 201 6 US-10-995-561-32124 Sequence 32124, A
C 133 16.8 66.4 201 6 US-10-995-561-32125 Sequence 32125, A
C 134 16.6 66.4 486 6 US-10-750-185-61106 Sequence 61106, A
C 135 16.6 66.4 600 6 US-10-750-185-2395 Sequence 2395, App
C 136 16.6 66.4 600 6 US-10-750-185-20950 Sequence 20950, A
C 137 16.6 66.4 1238 6 US-10-750-185-63868 Sequence 63868, A
C 138 16.6 66.4 1243 6 US-10-750-185-64558 Sequence 64558, A
C 139 16.6 66.4 1246 6 US-10-750-185-39113 Sequence 39113, A
C 140 16.6 66.4 1282 6 US-10-750-185-62341 Sequence 62341, A
C 141 16.6 66.4 1282 6 US-10-750-185-47195 Sequence 47195, A
C 142 16.6 66.4 1306 6 US-10-750-185-50239 Sequence 50239, A
C 143 16.6 66.4 1332 6 US-10-750-185-47384 Sequence 47384, A
C 144 16.6 66.4 1419 6 US-10-750-185-46948 Sequence 46948, A
C 145 16.6 66.4 1493 6 US-10-750-185-25125 Sequence 25125, A
C 146 16.6 66.4 1508 6 US-10-750-185-48745 Sequence 48745, A
C 147 16.6 66.4 1542 6 US-10-750-185-62797 Sequence 62797, A
C 148 16.6 66.4 1855 6 US-10-750-185-35000 Sequence 35000, A
C 149 16.6 66.4 1877 6 US-10-750-185-61205 Sequence 61205, A
C 150 16.6 66.4 1979 6 US-10-750-185-56295 Sequence 56295, A

ALIGNMENTS

RESULT 1
US-11-121-086-31
; Sequence 31, Application US/11121086
; Publication No. US2005026459A1
; GENERAL INFORMATION:
; APPLICANT: NIELSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107

SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31
; LENGTH: 218821
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: modified_base
; LOCATION: (106949)..(106949)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (110322)..(110324)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (115133)..(115133)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (131300)..(131300)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (139059)..(139158)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (157740)..(157740)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (157777)..(157777)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (157900)..(157900)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (157919)..(157919)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (158094)..(158094)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (158138)..(158138)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (158193)..(158195)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (158241)..(158242)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (158259)..(158259)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (158278)..(158278)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE: modified_base
; LOCATION: (158295)..(158295)
; OTHER INFORMATION: a, c, g, t, unknown or other

```
FEATURE:
NAME/KEY: modified_base
LOCATION: (158740)..(158839)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158929)..(158929)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163528)..(163530)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163550)..(163550)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163765)..(163765)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164000)..(164000)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164047)..(164047)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164084)..(164084)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167233)..(167233)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167236)..(167236)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167238)..(167238)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170247)..(170247)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170250)..(170250)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170252)..(170252)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170259)..(170259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170262)..(170262)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170266)..(170266)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174470)..(174470)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
```

```
NAME/KEY: modified_base
LOCATION: (174472)..(174472)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174474)..(174474)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179059)..(179060)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179064)..(179064)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197001)..(197001)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197004)..(197005)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197007)..(197007)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (200349)..(200349)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (200351)..(200351)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (212425)..(212426)
OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-31
```

```
Query Match          79.2%; Score 19.8; DB 7; Length 218821;
Best Local Similarity 91.3%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  AAAAAAAAAAGCATGCTGTGAC 23
Db      32967  AAAAAAAAAAGCATGCTGTGCGC 32989
```

```
RESULT 2
US-11-121-086-8
; Sequence 8, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 8
; LENGTH: 246960
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-8
```

```
Query Match          77.6%; Score 19.4; DB 7; Length 246960;
Best Local Similarity 95.2%; Pred. No. 52;
```

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGTG 21
Db 101580 AAAAAAAAAAGCATGCTGTG 101600

RESULT 3
US-10-750-185-618
; Sequence 618, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 618
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT07416
US-10-750-185-618

Query Match 76.8%; Score 19.2; DB 6; Length 600;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
Db 353 AAAAAAAAAAGCATCTCTGTGACA 376

RESULT 4
US-10-750-185-26597
; Sequence 26597, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26597
; LENGTH: 2998
; TYPE: DNA
; ORGANISM: Bovine 19866881169843
US-10-750-185-26597

Query Match 76.8%; Score 19.2; DB 6; Length 2998;
Best Local Similarity 87.5%; Pred. No. 33;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGTGACA 24

Db 1904 AAAAAAAAAAGCAATCTGTAAACA 1927

RESULT 5
US-10-750-185-27841
; Sequence 27841, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27841
; LENGTH: 1884
; TYPE: DNA
; ORGANISM: Bovine 19866880509629
US-10-750-185-27841

Query Match 75.2%; Score 18.8; DB 6; Length 1884;
Best Local Similarity 90.9%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGA 22
Db 367 AAAAAAAAAAGCAACTGTGA 388

RESULT 6
US-11-112-908-60
; Sequence 60, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 60
; LENGTH: 171427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-60

Query Match 74.4%; Score 18.6; DB 7; Length 171427;
Best Local Similarity 84.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGCTGTGACAC 25
Db 165790 AAAAAAAAAATCATGAAGTGTATAC 165814

RESULT 7
US-10-750-185-46229
; Sequence 46229, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46229
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine 19866881024060
US-10-750-185-46229

Query Match 73.6%; Score 18.4; DB 6; Length 600;
Best Local Similarity 95.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGACTGT 20
Db 581 AAAAAAAAAAGCATGACTGT 600

RESULT 8
US-10-750-185-20178/c
; Sequence 20178, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20178
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Bovine MM118831
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: n is any nucleotide
US-10-750-185-20178

Query Match 72.8%; Score 18.2; DB 6; Length 596;
Best Local Similarity 87.0%; Pred. No. 67;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGACTGTAC 23
|||||||

Db 80 AAAAAAAAAAGCAAGTTGTGAC 58
RESULT 9
US-10-750-185-39819/c
; Sequence 39819, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39819
; LENGTH: 875
; TYPE: DNA
; ORGANISM: Bovine 19866880826868
US-10-750-185-39819

Query Match 72.8%; Score 18.2; DB 6; Length 875;
Best Local Similarity 87.0%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGACTGTGAC 23
Db 359 AAAAAAAAAAGCAAGTTGTGAC 337

RESULT 10
US-10-995-561-49515/c
; Sequence 49515, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49515
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-49515

Query Match 71.2%; Score 17.8; DB 6; Length 201;
Best Local Similarity 90.5%; Pred. No. 83;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGACTGTG 21
Db 89 AAAAAAAAAAGCATGCTGTG 69

RESULT 11
US-10-995-561-49533/c
; Sequence 49533, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.

```
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: PaacSeq for Windows Version 4.0
/ SEQ ID NO 49533
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561.49533

Query Match          71.2%; Score 17.8; DB 6; Length 201;
Best Local Similarity 90.5%; Pred. No. 83;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
   |||||
Db 95 AAAAAAAAAAGCATGACTGTG 75

RESULT 12
US-10-750-185-26879/c
/ Sequence 26879, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 26879
/ LENGTH: 1311
/ TYPE: DNA
/ ORGANISM: Bovine 19866881611582
US-10-750-185-26879

Query Match          71.2%; Score 17.8; DB 6; Length 1311;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
   |||||
Db 685 AAAAAAAAAAGCATGCTATG 665

RESULT 13
US-10-750-185-34711/c
/ Sequence 34711, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
```

```
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 34711
/ LENGTH: 1490
/ TYPE: DNA
/ ORGANISM: Bovine 1986688065470
US-10-750-185-34711

Query Match          71.2%; Score 17.8; DB 6; Length 1490;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
   |||||
Db 271 AAAAAAAAAAGCATGACTGTG 251

RESULT 14
US-10-750-185-39332/c
/ Sequence 39332, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 39332
/ LENGTH: 2204
/ TYPE: DNA
/ ORGANISM: Bovine 19866880818887
US-10-750-185-39332

Query Match          71.2%; Score 17.8; DB 6; Length 2204;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTG 21
   |||||
Db 2053 AAAAAAAAAAGCATGACTGAG 2033

RESULT 15
US-10-995-561-13301/c
/ Sequence 13301, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: PaacSeq for Windows Version 4.0
/ SEQ ID NO 13301
/ LENGTH: 35997
/ TYPE: DNA
/ ORGANISM: Homo sapiens
```


FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(35997)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13301

Query Match 71.2%; Score 17.8; DB 6; Length 35997;
Best Local Similarity 90.5%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTG 21
DB 26751 AAAAAAAAAATAGCATGCTGTG 26731

RESULT 16
US-10-995-561-23252
Sequence 23252, Application US/10995561
Publication No. US20050272054A1

GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CU001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23252
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-23252

Query Match 70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
DB 150 AAAAAAAAAAGTGACTGTCTCA 173

RESULT 17
US-10-995-561-58952/c
Sequence 58952, Application US/10995561
Publication No. US20050272054A1

GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CU001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 58952
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-58952

Query Match 70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
DB 177 AAAAAAAAAAGATGCTGTGAGAA 154

RESULT 18
US-10-750-185-35652
Sequence 35652, Application US/10750185
Publication No. US20050260603A1

GENERAL INFORMATION:
APPLICANT: MMT GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35652
LENGTH: 737
TYPE: DNA
ORGANISM: Bovine 1986681468215
US-10-750-185-35652

Query Match 70.4%; Score 17.6; DB 6; Length 737;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
DB 351 AAAAAAAAAAGCAAAATTGTGAGA 374

RESULT 19
US-10-750-185-59865
Sequence 59865, Application US/10750185
Publication No. US20050260603A1

GENERAL INFORMATION:
APPLICANT: MMT GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 59865
LENGTH: 1107
TYPE: DNA
ORGANISM: Bovine 1986681350281
US-10-750-185-59865

Query Match 70.4%; Score 17.6; DB 6; Length 1107;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGCTGTGACA 24
DB 727 AAGAAAATAAGCATGACCTGTGACA 750

RESULT 20
US-10-750-185-30513/c
Sequence 30513, Application US/10750185

```

; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30513
; LENGTH: 1328
; TYPE: DNA
; ORGANISM: Bovine 19866881384544
US-10-750-185-30513

Query Match          70.4%; Score 17.6; DB 6; Length 1328;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 422 AAAAAAAAAAGCATGAAAGACACA 399

RESULT 21
US-10-750-185-60402/c
; Sequence 60402, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60402
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Bovine 19866880874524
US-10-750-185-60402

Query Match          70.4%; Score 17.6; DB 6; Length 1353;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 473 AAACAAGAGCAACTGTGACA 450

RESULT 22
US-10-750-185-54530
; Sequence 54530, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.

```

```

; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54530
; LENGTH: 1813
; TYPE: DNA
; ORGANISM: Bovine 19866881039934
US-10-750-185-54530

Query Match          70.4%; Score 17.6; DB 6; Length 1813;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 409 AAAAAAAAAAGCAAGACTATGAAA 432

RESULT 23
US-10-750-185-61429/c
; Sequence 61429, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61429
; LENGTH: 2275
; TYPE: DNA
; ORGANISM: Bovine 19866880595799
US-10-750-185-61429

Query Match          70.4%; Score 17.6; DB 6; Length 2275;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
Db 362 AAAAAAAAAAGCATGTTCTTGACA 339

RESULT 24
US-10-750-185-39842/c
; Sequence 39842, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David

```

```
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: NM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39842
; LENGTH: 2927
; TYPE: DNA
; ORGANISM: Bovine 19866880775971
US-10-750-185-39842

Query Match          70.4%; Score 17.6; DB 6; Length 2927;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
    |||||
Db 148 AAAAAAAAAAGCATGACTGTGACA 125

RESULT 25
US-10-995-561-13341/c
; Sequence 13341, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13341
; LENGTH: 28693
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13341

Query Match          70.4%; Score 17.6; DB 6; Length 28693;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
    |||||
Db 22787 AAAAAAAAAAGCATGACTGTGAAA 22764

RESULT 26
US-10-995-561-13254/c
; Sequence 13254, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13254
; LENGTH: 65931
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-10-995-561-13254

Query Match          70.4%; Score 17.6; DB 6; Length 65931;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
    |||||
Db 35118 AAAAAAAAAAGCATGACTGTGACCA 35095

RESULT 27
US-10-857-780-2/c
; Sequence 2, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOVAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 110950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-857-780-2

Query Match          70.4%; Score 17.6; DB 6; Length 110950;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
    |||||
Db 9726 AAAAAAAAAAGCATGACTGTGCA 9703

RESULT 28
US-11-112-908-35/c
; Sequence 35, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: HARRIS, Cole
; APPLICANT: DAVIS, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 127340
```

```
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-112-908-35

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 7; Length 127340;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 125464 AAAAAAAAAACAGATGGAGACA 125441

RESULT 29
US-11-112-908-64
/ Sequence 64, Application US/11112908
/ Publication No. US20050260659A1
/ GENERAL INFORMATION:
/ APPLICANT: Harris, Cole
/ TITLE OF INVENTION: Breast Cancer Biomarkers
/ FILE REFERENCE: 04-164-US
/ CURRENT APPLICATION NUMBER: US/11/112,908
/ PRIOR FILING DATE: 2005-04-22
/ PRIOR APPLICATION NUMBER: US 60/564,758
/ PRIOR FILING DATE: 2004-04-23
/ PRIOR APPLICATION NUMBER: US 60/575,978
/ PRIOR FILING DATE: 2004-06-01
/ PRIOR APPLICATION NUMBER: US 60/631,702
/ PRIOR FILING DATE: 2004-11-30
/ PRIOR APPLICATION NUMBER: US 60/633,826
/ PRIOR FILING DATE: 2004-12-07
/ NUMBER OF SEQ ID NOS: 511
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 64
/ LENGTH: 157230
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-112-908-64

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 7; Length 157230;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 149845 AAAAAAAAAAGCCTGATCTTACA 149868

RESULT 30
US-11-112-908-62
/ Sequence 62, Application US/11112908
/ Publication No. US20050260659A1
/ GENERAL INFORMATION:
/ APPLICANT: Harris, Cole
/ TITLE OF INVENTION: Breast Cancer Biomarkers
/ FILE REFERENCE: 04-164-US
/ CURRENT APPLICATION NUMBER: US/11/112,908
/ PRIOR FILING DATE: 2005-04-22
/ PRIOR APPLICATION NUMBER: US 60/564,758
/ PRIOR FILING DATE: 2004-04-23
/ PRIOR APPLICATION NUMBER: US 60/575,978
/ PRIOR FILING DATE: 2004-06-01
/ PRIOR APPLICATION NUMBER: US 60/631,702
/ PRIOR FILING DATE: 2004-11-30
/ PRIOR APPLICATION NUMBER: US 60/633,826
/ PRIOR FILING DATE: 2004-12-07
/ NUMBER OF SEQ ID NOS: 511
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 62
/ LENGTH: 170508
/ TYPE: DNA

/ ORGANISM: Homo sapiens
US-11-112-908-62

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 7; Length 170508;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 133601 AAAAAAAAAAGCCTGATCTTACA 133624

RESULT 31
US-11-112-908-65
/ Sequence 65, Application US/11112908
/ Publication No. US20050260659A1
/ GENERAL INFORMATION:
/ APPLICANT: Harris, Cole
/ TITLE OF INVENTION: Breast Cancer Biomarkers
/ FILE REFERENCE: 04-164-US
/ CURRENT APPLICATION NUMBER: US/11/112,908
/ PRIOR FILING DATE: 2005-04-22
/ PRIOR APPLICATION NUMBER: US 60/564,758
/ PRIOR FILING DATE: 2004-04-23
/ PRIOR APPLICATION NUMBER: US 60/575,978
/ PRIOR FILING DATE: 2004-06-01
/ PRIOR APPLICATION NUMBER: US 60/631,702
/ PRIOR FILING DATE: 2004-11-30
/ PRIOR APPLICATION NUMBER: US 60/633,826
/ PRIOR FILING DATE: 2004-12-07
/ NUMBER OF SEQ ID NOS: 511
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 65
/ LENGTH: 173115
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-112-908-65

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 7; Length 173115;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGACA 24
DB 85356 AAAAAAAAAAGCCTGATCTTACA 85379

RESULT 32
US-11-121-086-55
/ Sequence 55, Application US/11121086
/ Publication No. US2005026459A1
/ GENERAL INFORMATION:
/ APPLICANT: NIELSEN, KIRSTEN V.
/ TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
/ FILE REFERENCE: 09138.6000-00000
/ CURRENT APPLICATION NUMBER: US/11/121,086
/ PRIOR FILING DATE: 2005-05-04
/ PRIOR APPLICATION NUMBER: 60/567,570
/ PRIOR FILING DATE: 2004-05-04
/ NUMBER OF SEQ ID NOS: 107
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 55
/ LENGTH: 175673
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-11-121-086-55

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 7; Length 175673;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

OY 1 AAAAAAAAAAGCATGCTGTGACA 24
|||||
Db 48166 AAAAAAAAAATCATTAAGTGTGACA 48189

RESULT 33
US-11-112-908-32/c
; Sequence 32, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 32
; LENGTH: 193363
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-32

Query Match 70.4%; Score 17.6; DB 7; Length 193363;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGCTGTGACA 24
|||||
Db 47881 AAAAAAAAAAATCACTGTTAA 47858

RESULT 34
US-10-995-561-13227
; Sequence 13227, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13227
; LENGTH: 317876
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(317876)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13227

Query Match 70.4%; Score 17.6; DB 6; Length 317876;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGCTGTGACA 24
|||||
Db 97860 AAAAAAAAAAGGATGCTGTGCA 97883

RESULT 35
US-10-995-561-13293
; Sequence 13293, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13293
; LENGTH: 645179
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13293

Query Match 70.4%; Score 17.6; DB 6; Length 645179;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGCTGTGACA 24
|||||
Db 307897 AAAAAACCAAGCATGCTGAGAA 307920

RESULT 36
US-11-121-849-24059/c
; Sequence 24059, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 24059
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-24059

Query Match 69.6%; Score 17.4; DB 7; Length 25;
Best Local Similarity 94.7%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 7 AAAAAAAAAAGCATGCTGTGACAC 25
|||||
Db 23 AAAAAATGATGTGTGACAC 5

RESULT 37
US-11-121-849-454515/c
; Sequence 454515, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Method of Genetic Analysis of Formalin Fixed Paraffin Embedded
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03

```

; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 454515
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-049-454515

Query Match
Best Local Similarity 69.6%; Score 17.4; DB 7; Length 25;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
7 AAAAGCATGCTGTGACAC 25
22 AAAAGCATGATGTGACAC 4

Db
22 AAAAGCATGATGTGACAC 4

RESULT 38
US-11-121-086-7/C
; Sequence 7, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7
; LENGTH: 162085
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-7

Query Match
Best Local Similarity 69.6%; Score 17.4; DB 7; Length 162085;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGCATGACTG 19
93914 AAAAAAAAAAGATGACTG 93896

Db
93914 AAAAAAAAAAGATGACTG 93896

RESULT 39
US-11-121-086-43
; Sequence 43, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 43
; LENGTH: 175416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-43

Query Match
Best Local Similarity 69.6%; Score 17.4; DB 7; Length 175416;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGCATGACTG 19
AAAAAAAAAAGCATGACTG 19
```

```

Db
36425 AAAAAAAAAAGATGACTG 36443

RESULT 40
US-10-995-561-49228/C
; Sequence 49228, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49228
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-49228

Query Match
Best Local Similarity 68.8%; Score 17.2; DB 6; Length 201;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGCATGCTGACA 24
111 AAAAAAAAAAGATCACTTTGACA 88

Db
111 AAAAAAAAAAGATCACTTTGACA 88

RESULT 41
US-10-995-561-54497/C
; Sequence 54497, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54497
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-54497

Query Match
Best Local Similarity 68.8%; Score 17.2; DB 6; Length 201;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGCATGCTGCA 22
99 AAAAAAAAAAGATGACTGTCA 78

Db
99 AAAAAAAAAAGATGACTGTCA 78

RESULT 42
US-10-995-561-54499/C
; Sequence 54499, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
```

CURRENT FILING DATE: 2004-11-24
 NUMBER OF SEQ ID NOS: 85702
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 54499
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-995-561-54499

Query Match 68.8%; Score 17.2; DB 6; Length 201;
 Best Local Similarity 86.4%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
 |||||
 Db 100 AAAAAAAAAAGCATGACTGTGA 79

RESULT 43
 US-10-995-561-55548/C
 Sequence 55548, Application US/10995561
 Publication No. US20050272054A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michele et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 FILE REFERENCE: C1001559
 CURRENT APPLICATION NUMBER: US/10/995,561
 CURRENT FILING DATE: 2004-11-24
 NUMBER OF SEQ ID NOS: 85702
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 55548
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-995-561-55548

Query Match 68.8%; Score 17.2; DB 6; Length 201;
 Best Local Similarity 86.4%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
 |||||
 Db 138 AAAAAAAAAAGCATGACTGTGA 117

RESULT 44
 US-10-995-561-55551/C
 Sequence 55551, Application US/10995561
 Publication No. US20050272054A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michele et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 FILE REFERENCE: C1001559
 CURRENT APPLICATION NUMBER: US/10/995,561
 CURRENT FILING DATE: 2004-11-24
 NUMBER OF SEQ ID NOS: 85702
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 55551
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-995-561-55551

Query Match 68.8%; Score 17.2; DB 6; Length 201;
 Best Local Similarity 86.4%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGA 22
 |||||

Db 149 AAAAAAAAAAGCATGACTGTGA 128

RESULT 45
 US-10-750-185-20253
 Sequence 20253, Application US/10750185
 Publication No. US20050260603A1
 GENERAL INFORMATION:
 APPLICANT: MMI GENOMICS, INC.
 APPLICANT: DENISE, Sue K.
 APPLICANT: KERR, Richard
 APPLICANT: ROSENFELD, David
 APPLICANT: HOLM, Tom
 APPLICANT: BATES, Stephen
 APPLICANT: FANTIN, Dennis
 TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 FILE REFERENCE: MM1100-2
 CURRENT APPLICATION NUMBER: US/10/750,185
 CURRENT FILING DATE: 2003-12-31
 PRIOR APPLICATION NUMBER: US 60/437,482
 PRIOR FILING DATE: 2002-12-31
 NUMBER OF SEQ ID NOS: 64922
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 20253
 LENGTH: 600
 TYPE: DNA
 ORGANISM: Bovine
 US-10-750-185-20253

Query Match 68.8%; Score 17.2; DB 6; Length 600;
 Best Local Similarity 86.4%; Pred. No. 1.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
 |||||
 Db 135 AAAAAAAAAAGCATGACTGTGA 156

RESULT 46
 US-10-750-185-45897/C
 Sequence 45897, Application US/10750185
 Publication No. US20050260603A1
 GENERAL INFORMATION:
 APPLICANT: MMI GENOMICS, INC.
 APPLICANT: DENISE, Sue K.
 APPLICANT: KERR, Richard
 APPLICANT: ROSENFELD, David
 APPLICANT: HOLM, Tom
 APPLICANT: BATES, Stephen
 APPLICANT: FANTIN, Dennis
 TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 FILE REFERENCE: MM1100-2
 CURRENT APPLICATION NUMBER: US/10/750,185
 CURRENT FILING DATE: 2003-12-31
 PRIOR APPLICATION NUMBER: US 60/437,482
 PRIOR FILING DATE: 2002-12-31
 NUMBER OF SEQ ID NOS: 64922
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 45897
 LENGTH: 997
 TYPE: DNA
 ORGANISM: Bovine
 US-10-750-185-45897

Query Match 68.8%; Score 17.2; DB 6; Length 997;
 Best Local Similarity 86.4%; Pred. No. 1.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGACTGTGA 22
 |||||
 Db 635 AAAAAAAAAAGCATGACTGTGA 614

```
RESULT 47
US-10-750-185-30612/c
; Sequence 30612, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30612
; LENGTH: 1025
; TYPE: DNA
; ORGANISM: Bovine 19866882114681
US-10-750-185-30612

Query Match      68.8%; Score 17.2; DB 6; Length 1025;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
Db 470 AAAAAAAAAAGCATGTACAGTAA 449

RESULT 48
US-10-750-185-40180
; Sequence 40180, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40180
; LENGTH: 1202
; TYPE: DNA
; ORGANISM: Bovine 19866880710439
US-10-750-185-40180

Query Match      68.8%; Score 17.2; DB 6; Length 1202;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
Db 249 AAAAAAAAAAGCATGACATTGA 270
```

```
RESULT 49
US-10-750-185-51004/c
; Sequence 51004, Application US/10750185
```

```
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51004
; LENGTH: 1299
; TYPE: DNA
; ORGANISM: Bovine 19866881183947
US-10-750-185-51004

Query Match      68.8%; Score 17.2; DB 6; Length 1299;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGACTGTGA 22
Db 1274 AAAAAAAAAATCATGACTTTTA 1253
```

```
RESULT 50
US-10-750-185-26061/c
; Sequence 26061, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26061
; LENGTH: 1444
; TYPE: DNA
; ORGANISM: Bovine 19866880567251
US-10-750-185-26061

Query Match      68.8%; Score 17.2; DB 6; Length 1444;
Best Local Similarity 86.4%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGCATGACTGTGACA 24
Db 973 AAAAAATPAAGCATGACATTGAAA 952
```

Search completed: December 14, 2005, 11:40:31
Job time : 187.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-6

Perfect score: 25

Sequence: 1 aaaaaaaaaagcatgatgtgtacac 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sta:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hlg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	41824	14	AC068427 Homo sapi
2	25	100.0	54810	14	AC027543 Homo sapi
3	25	100.0	153267	14	AC011212 Homo sapi
4	25	100.0	167272	8	AC011466 Homo sapi
5	22.4	89.6	50277	14	AC138524_5
6	22.4	89.6	110000	14	AC138524_4
7	22.4	89.6	116673	8	AC093755 Homo sapi
8	21.8	87.2	168056	14	AC125167 Mus muscu
9	21.8	87.2	181432	14	AC134962 Sus scrofa
10	21.8	87.2	211153	14	AC103436 Rattus no
11	21.8	87.2	300132	14	AC098535 Rattus no
12	21.4	85.6	255285	14	AC111222 Rattus no
13	21	84.0	32404	6	C0868604 Sequence
14	21	84.0	32404	6	C0868604 Sequence
15	20.8	83.2	2013	6	AX506824 Sequence
16	20.8	83.2	2013	6	AX506824 Sequence
17	20.8	83.2	51202	5	BX511033 Zebrafish
18	20.8	83.2	96729	14	CT009634 Danio rer

19	20.8	83.2	107527	15	P19P19	AC000104 Sequence
20	20.8	83.2	110000	14	AC094469_0	AC094469 Rattus no
21	20.8	83.2	110000	14	BX308758_1	Continuation (2 of
22	20.8	83.2	124127	15	AC149546	AC149546 Medicago
23	20.8	83.2	144093	9	AC109193	AC109193 Mus muscu
24	20.8	83.2	144375	14	CR354586	CR354586 Danio rer
25	20.8	83.2	144639	15	AC138453	AC138453 Medicago
26	20.8	83.2	149414	5	BX957226	BX957226 Zebrafish
27	20.8	83.2	170941	8	AL606970	AL606970 Human DNA
28	20.8	83.2	170941	8	AC069536	AC069536 Homo sapi
29	20.8	83.2	172823	14	AC018740	AC018740 Homo sapi
30	20.8	83.2	172969	8	AC090574	AC090574 Homo sapi
31	20.8	83.2	180450	2	AE014835	AE014835 Plasmodu
32	20.8	83.2	181137	14	AC084303	AC084303 Homo sapi
33	20.8	83.2	181400	8	AC084251	AC084251 Homo sapi
34	20.8	83.2	186014	9	AC133942	AC133942 Mus muscu
35	20.8	83.2	191146	5	BX530070	BX530070 Zebrafish
36	20.8	83.2	197946	9	AC098722	AC098722 Mus muscu
37	20.8	83.2	201611	14	AC021185	AC021185 Homo sapi
38	20.8	83.2	202495	14	AC149854	AC149854 Papio anu
39	20.8	83.2	203281	9	AC108435	AC108435 Mus muscu
40	20.8	83.2	205761	5	BX510655	BX510655 Zebrafish
41	20.8	83.2	208833	14	AC161795	AC161795 Mus muscu
42	20.8	83.2	213024	14	AC163593	AC163593 Bos tauru
43	20.8	83.2	213666	14	AC161925	AC161925 Mus muscu
44	20.8	83.2	217167	9	AC122327	AC122327 Mus muscu
45	20.8	83.2	220189	14	AC094406	AC094406 Rattus no
46	20.8	83.2	224364	14	AC094468	AC094468 Rattus no
47	20.8	83.2	225581	14	BX537105	BX537105 Danio rer
48	20.8	83.2	231059	14	AC165350	AC165350 Mus muscu
49	20.8	83.2	250743	2	AE014836	AE014836 Plasmodu
50	20.8	83.2	263479	14	AC125986	AC125986 Rattus no
51	20.4	81.6	110000	14	PFMALBP1_08	Continuation (9 of
52	20.4	81.6	169218	14	AC149013	AC149013 Papio anu
53	20.4	81.6	195462	5	BX004776	BX004776 Zebrafish
54	20.4	81.6	198108	5	BX537104	BX537104 Zebrafish
55	20.4	81.6	199660	14	AC099610	AC099610 Mus muscu
56	20.4	81.6	213247	14	CR388139	CR388139 Danio rer
57	20.4	81.6	230161	9	AL772135	AL772135 Mouse DNA
58	20.2	80.8	261	6	AR556408	AR556408 Sequence
59	20.2	80.8	726	10	BV564074	BV564074 G591P6139
60	20.2	80.8	801	10	BV580970	BV580970 G591P6139
61	20.2	80.8	946	10	BV570112	BV570112 G591P6056
62	20.2	80.8	6861	2	AY326456	AY326456 Spodoptex
63	20.2	80.8	6861	2	AX659386	AX659386 Sequence
64	20.2	80.8	39131	4	AC155900	AC155900 Toxodonta
65	20.2	80.8	62729	14	AC100913	AC100913 Mus muscu
66	20.2	80.8	62729	14	AC100913	AC100913 Mus muscu
67	20.2	80.8	73211	14	AC166462	AC166462 Bos tauru
68	20.2	80.8	86351	9	AL929244	AL929244 Mouse DNA
69	20.2	80.8	92850	5	BX284628	BX284628 Zebrafish
70	20.2	80.8	94606	5	BX539345	BX539345 Zebrafish
71	20.2	80.8	95185	15	ATP617	AL049657 Arabidops
72	20.2	80.8	95618	5	BX569801	BX569801 Zebrafish
73	20.2	80.8	99688	15	ATF13M23	AL035523 Arabidops
74	20.2	80.8	108614	14	AC137100	AC137100 Homo sapi
75	20.2	80.8	109786	15	FSK24	AL28395 Arabidops
76	20.2	80.8	110000	15	AP008208_090	Continuation (90 o
77	20.2	80.8	110000	15	AP008208_090	Continuation (90 o
78	20.2	80.8	110000	15	AP008210_306	Continuation (91 o
79	20.2	80.8	111200	8	AC093227	AC093227 Homo sapi
80	20.2	80.8	112621	8	AC004891	AC004891 Homo sapi
81	20.2	80.8	117951	8	AL391803	AL391803 Human DNA
82	20.2	80.8	131189	14	AC101756	AC101756 Mus muscu
83	20.2	80.8	131313	14	AC108902	AC108902 Felis cat
84	20.2	80.8	132715	5	BX465189	BX465189 Zebrafish
85	20.2	80.8	133263	14	AC126235	AC126235 Canis fam
86	20.2	80.8	133540	14	AC152048	AC152048 Daaypus n
87	20.2	80.8	133568	8	HS16915	AL732589 Mouse DNA
88	20.2	80.8	134644	9	AL732589	AL732589 Mouse DNA
89	20.2	80.8	136939	14	AC149521	AC149521 Xenopus t
90	20.2	80.8	137366	14	AB003979	AB003979 Oryza bat
91	20.2	80.8	140479	15	AC157893	AC157893 Medicago


```

* 7309 7408: gap of 100 bp
* 7409 8080: contig of 672 bp in length
* 8081 8180: gap of 100 bp
* 8181 8911: contig of 731 bp in length
* 8912 9011: gap of 100 bp
* 9012 9741: contig of 730 bp in length
* 9742 9841: gap of 100 bp
* 9842 10565: contig of 724 bp in length
* 10566 10665: gap of 100 bp
* 10666 11404: contig of 739 bp in length
* 11405 11505: gap of 100 bp
* 11506 12208: contig of 704 bp in length
* 12209 12308: gap of 100 bp
* 12309 13034: contig of 726 bp in length
* 13035 13134: gap of 100 bp
* 13135 13853: contig of 719 bp in length
* 13854 13953: gap of 100 bp
* 13954 14699: contig of 746 bp in length
* 14700 14799: gap of 100 bp
* 14800 15523: contig of 724 bp in length
* 15524 15623: gap of 100 bp
* 15624 16392: contig of 769 bp in length
* 16393 16492: gap of 100 bp
* 16493 17224: contig of 732 bp in length
* 17225 17324: gap of 100 bp
* 17325 18048: contig of 724 bp in length
* 18049 18148: gap of 100 bp
* 18149 18877: contig of 729 bp in length
* 18878 18977: gap of 100 bp
* 18978 19710: contig of 733 bp in length
* 19711 19810: gap of 100 bp
* 19811 20549: contig of 735 bp in length
* 20550 20649: gap of 100 bp
* 20650 21361: contig of 712 bp in length
* 21362 21461: gap of 100 bp
* 21462 22168: contig of 707 bp in length
* 22169 22268: gap of 100 bp
* 22269 22971: contig of 703 bp in length
* 22972 23071: gap of 100 bp
* 23072 23803: contig of 732 bp in length
* 23804 23903: gap of 100 bp
* 23904 24579: contig of 676 bp in length
* 24580 24679: gap of 100 bp
* 24680 25406: contig of 727 bp in length
* 25407 25506: gap of 100 bp
* 25507 26229: contig of 723 bp in length
* 26230 26329: gap of 100 bp
* 26330 27067: contig of 738 bp in length
* 27068 27167: gap of 100 bp
* 27168 27888: contig of 721 bp in length
* 27889 27988: gap of 100 bp
* 27989 28687: contig of 699 bp in length
* 28688 28787: gap of 100 bp
* 28788 29500: contig of 713 bp in length
* 29501 29600: gap of 100 bp
* 29601 30277: contig of 677 bp in length
* 30278 30377: gap of 100 bp
* 30378 31110: contig of 733 bp in length
* 31111 31210: gap of 100 bp
* 31211 31938: contig of 728 bp in length
* 31939 32038: gap of 100 bp
* 32039 32752: contig of 714 bp in length
* 32753 32852: gap of 100 bp
* 32853 33572: contig of 720 bp in length
* 33573 33672: gap of 100 bp
* 33673 34391: contig of 719 bp in length
* 34392 34491: gap of 100 bp
* 34492 35224: contig of 733 bp in length
* 35225 35324: gap of 100 bp
* 35325 36054: contig of 730 bp in length
* 36055 36154: gap of 100 bp
* 36155 36883: contig of 729 bp in length
* 36884 36983: gap of 100 bp

```

```

* 36984 37701: contig of 718 bp in length
* 37702 37801: gap of 100 bp
* 37802 38512: contig of 711 bp in length
* 38513 38612: gap of 100 bp
* 38613 39340: contig of 728 bp in length
* 39341 39441: gap of 100 bp
* 39441 40178: contig of 738 bp in length
* 40179 40278: gap of 100 bp
* 40279 41007: contig of 729 bp in length
* 41008 41107: gap of 100 bp
* 41108 41824: contig of 717 bp in length.

FEATURES
source
1..41824
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12"
/clone="RP11-252G6"
/clone_11b="RP11 Human Male BAC"
726..825
/estimated_length=100
1551..1650
/estimated_length=100
2372..2471
/estimated_length=100
3196..3295
/estimated_length=100
4014..4113
/estimated_length=100
4839..4938
/estimated_length=100
5661..5760
/estimated_length=100
6489..6588
/estimated_length=100
7309..7408
/estimated_length=100
8081..8180
/estimated_length=100
8912..9011
/estimated_length=100
9742..9841
/estimated_length=100
10566..10665
/estimated_length=100
11405..11504

Query Match 100.0%; Score 25; DB 14; Length 41824;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATGTGCAC 25
Db 39148 AAAAAAAAAAGCATGATGTGCAC 39124

RESULT 2
AC027543 54810 bp DNA linear HTG 28-AUG-2002
AC027543
SEQUENCE SAMPLING.
AC027543.1 GI:7342288
AC027543.1 GI:7342288
HTG: HTGS PHASRO.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 54810)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 19, clone RP11-601G24
TITLE
Homo sapiens chromosome 19, clone RP11-601G24

```

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 54810)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campianno,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Labocque,K., Lamazares,R., Landers,T., Lebecky,J.,
Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

JOURNAL
REFERENCE
AUTHORS
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 54810)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campianno,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Labocque,K., Lamazares,R., Landers,T., Lebecky,J.,
Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

JOURNAL
REFERENCE
AUTHORS
Submitted (28-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8832
Center clone name: 601_G_24

* NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

* be preserved.
* 1
* 708: contig of 708 bp in length
* 709
* 808: gap of 100 bp
* 809
* 1522: contig of 714 bp in length
* 1523
* 1522: gap of 100 bp
* 1623
* 2290: contig of 668 bp in length
* 2291
* 2290: gap of 100 bp
* 2391
* 3078: contig of 688 bp in length
* 3079
* 3178: gap of 100 bp
* 3179
* 3873: contig of 695 bp in length
* 3874
* 3973: gap of 100 bp
* 3974
* 4659: contig of 696 bp in length
* 4769: gap of 100 bp
* 4770
* 5471: contig of 702 bp in length
* 5472
* 5571: gap of 100 bp
* 5572
* 6271: contig of 700 bp in length
* 6272
* 6371: gap of 100 bp
* 6372
* 7023: contig of 652 bp in length
* 7024
* 7123: gap of 100 bp
* 7124
* 7838: contig of 715 bp in length
* 7839
* 7938: gap of 100 bp
* 7939
* 8624: contig of 686 bp in length
* 8625
* 8724: gap of 100 bp
* 8725
* 9392: contig of 668 bp in length
* 9393
* 9492: gap of 100 bp
* 9493
* 10198: contig of 706 bp in length
* 10199
* 10298: gap of 100 bp
* 10299
* 10102: contig of 714 bp in length
* 11013
* 11112: gap of 100 bp
* 11113
* 11807: contig of 695 bp in length
* 11808
* 11907: gap of 100 bp
* 11908
* 12611: contig of 704 bp in length
* 12612
* 12711: gap of 100 bp
* 12712
* 13419: contig of 708 bp in length
* 13419
* 13519: gap of 100 bp
* 13520
* 14187: contig of 668 bp in length
* 14188
* 14287: gap of 100 bp
* 14288
* 14988: contig of 711 bp in length
* 14989
* 15098: gap of 100 bp
* 15099
* 15088: contig of 710 bp in length
* 15099
* 15908: gap of 100 bp
* 15909
* 15987: contig of 679 bp in length
* 15988
* 16587: gap of 100 bp
* 16588
* 17374: contig of 687 bp in length
* 17375
* 17474: gap of 100 bp
* 17475
* 18167: contig of 693 bp in length
* 18168
* 18267: gap of 100 bp
* 18268
* 18970: contig of 703 bp in length
* 18971
* 19070: gap of 100 bp
* 19071
* 19782: contig of 712 bp in length
* 19783
* 19882: gap of 100 bp
* 19883
* 20528: contig of 646 bp in length
* 20529
* 20628: gap of 100 bp
* 20629
* 21331: contig of 703 bp in length
* 21332
* 21431: gap of 100 bp
* 21432
* 22119: contig of 688 bp in length
* 22120
* 22119: gap of 100 bp
* 22220
* 22913: contig of 694 bp in length
* 22914
* 23013: gap of 100 bp
* 23014
* 23674: contig of 661 bp in length
* 23675
* 23774: gap of 100 bp
* 23775
* 24470: contig of 696 bp in length
* 24471
* 24570: gap of 100 bp
* 24571
* 25267: contig of 697 bp in length
* 25268
* 25367: gap of 100 bp
* 25368
* 26073: contig of 706 bp in length
* 26074
* 26173: gap of 100 bp
* 26174
* 26876: contig of 703 bp in length
* 26877
* 26976: gap of 100 bp
* 26977
* 27662: contig of 686 bp in length
* 27663
* 27762: gap of 100 bp
* 27763
* 28438: contig of 676 bp in length
* 28439
* 28439: gap of 100 bp

```

* 28539 29267: contig of 729 bp in length
* 29268 29367: gap of 100 bp
* 29368 30076: contig of 709 bp in length
* 30076 30176: gap of 100 bp
* 30176 30833: contig of 656 bp in length
* 30833 30933: gap of 100 bp
* 30933 31600: contig of 667 bp in length
* 31600 32390: gap of 100 bp
* 32390 32491: gap of 100 bp
* 32491 33189: contig of 698 bp in length
* 33189 33289: gap of 100 bp
* 33289 34102: contig of 714 bp in length
* 34102 34103: gap of 100 bp
* 34103 34801: contig of 699 bp in length
* 34801 34901: gap of 100 bp
* 34901 35595: contig of 694 bp in length
* 35595 35696: gap of 100 bp
* 35696 36353: contig of 657 bp in length
* 36353 36453: gap of 100 bp
* 36453 37251: contig of 698 bp in length
* 37251 37956: contig of 706 bp in length
* 37956 38057: gap of 100 bp
* 38057 38755: contig of 698 bp in length
* 38755 38854: gap of 100 bp
* 38854 39558: contig of 704 bp in length
* 39558 39659: gap of 100 bp
* 39659 40379: contig of 721 bp in length
* 40379 40479: gap of 100 bp
* 40479 41195: contig of 716 bp in length
* 41195 41296: gap of 100 bp
* 41296 41996: contig of 701 bp in length
* 41996 42097: gap of 100 bp
* 42097 42801: contig of 705 bp in length
* 42801 42901: gap of 100 bp
* 42901 43614: contig of 713 bp in length
* 43614 43715: gap of 100 bp
* 43715 44422: contig of 708 bp in length
* 44422 44522: gap of 100 bp

```

```

Query Match 100.0%; Score 25; DB 14; Length 54810;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
Db 10627 AAAAAAAAAAGCATGATTGTGACAC 10651

```

```

RESULT 3
AC011212/c 153267 bp DNA 1linear HTG 12-MAR-2000
LOCUS Homo sapiens clone RP11-3N16, WORKING DRAFT SEQUENCE, 22 unordered
DEFINITION pieces.
AC011212
AC011212 GI:7230012
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 153267)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-3N16
2 (bases 1 to 153267)
unpublished
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Bouhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,

```

TITLE JOURNAL COMMENT

```

Galagan,J., Gardyna,S., Grant,C., Hages,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lecoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talmas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced g1:6139137.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2783
Center clone name: 3_N_16
----- Summary Statistics
Sequencing vector: MJ3; M77815, 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 118139 bases at least Q40
Consensus quality: 133540 bases at least Q30
Consensus quality: 144810 bases at least Q20
Insert size: 15167; sum-of-contrigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1253 1252: contig of 1252 bp in length
* 1352 1352: gap of 100 bp
* 1353 2550: contig of 1198 bp in length
* 2551 2650: gap of 100 bp
* 2651 4705: contig of 2055 bp in length
* 4706 4805: gap of 100 bp
* 4806 7966: contig of 3161 bp in length
* 7967 8066: gap of 100 bp
* 8067 10438: contig of 2372 bp in length
* 10439 10538: gap of 100 bp
* 10539 13124: contig of 2566 bp in length
* 13125 13224: gap of 100 bp
* 13225 15586: contig of 2362 bp in length
* 15587 15686: gap of 100 bp
* 15687 19054: contig of 3368 bp in length
* 19055 19154: gap of 100 bp
* 19155 23442: contig of 4288 bp in length
* 23443 23542: gap of 100 bp
* 23543 28034: contig of 4492 bp in length
* 28035 28134: gap of 100 bp
* 28135 31029: contig of 2855 bp in length
* 31030 31129: gap of 100 bp
* 31130 36025: contig of 4836 bp in length
* 36026 36125: gap of 100 bp
* 36126 42963: contig of 6838 bp in length
* 42964 43063: gap of 100 bp
* 43064 51794: contig of 8731 bp in length
* 51795 51894: gap of 100 bp
* 51895 60991: contig of 9097 bp in length

```

```

* 60992 61091: gap of 100 bp
* 61092 70105: contig of 9014 bp in length
* 70106 70205: gap of 100 bp
* 70206 77517: contig of 7312 bp in length
* 77518 77617: gap of 100 bp
* 77618 87726: contig of 10109 bp in length
* 87727 87826: gap of 100 bp
* 87827 100723: contig of 12897 bp in length
* 100724 100823: gap of 100 bp
* 100824 115535: contig of 14712 bp in length
* 115536 115635: gap of 100 bp
* 115636 134718: contig of 19083 bp in length
* 134719 153267: contig of 18449 bp in length.
* 153267 Location/Qualifiers
  source          1..153267
                  /organism="Homo sapiens"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9606"
                  /clone_lib="RP11-3N16"
                  /clone_id="RP11-3N16 Human Male BAC"
  misc_feature    1..1252
                  /note="assembly_fragment"
  gap            1253..1352
                  /estimated_length=100
  misc_feature    1353..2550
                  /note="assembly_fragment"
  gap            2551..2650
                  /estimated_length=100
  misc_feature    2651..4705
                  /note="assembly_fragment"
  gap            4706..4805
                  /estimated_length=100
  misc_feature    4806..7966
                  /note="assembly_fragment"
  gap            7967..8066
                  /estimated_length=100
  misc_feature    8067..10438
                  /note="assembly_fragment"
  gap            10439..10538
                  /estimated_length=100
  misc_feature    10539..13124
                  /note="assembly_fragment"
  gap            13125..13224
                  /estimated_length=100
  misc_feature    13225..15586
                  /note="assembly_fragment"
  gap            15587..15686
                  /estimated_length=100
  misc_feature    15687..19054
                  /note="assembly_fragment"
  gap            19055..19154
                  /estimated_length=100
  misc_feature    19155..23442
                  /note="assembly_fragment"
  gap            23443..23542
                  /estimated_length=100
  misc_feature    23543..28034
                  /note="assembly_fragment"
  gap            28035..28134
                  /estimated_length=100
  misc_feature    28135..31029
                  /note="assembly_fragment"
  gap            31030..31129
                  /estimated_length=100
  misc_feature    31130..36025
                  /note="assembly_fragment"
  gap            36026..36125
                  /estimated_length=100
  misc_feature    36126..42963
                  /note="assembly_fragment"
  gap            42964..43063
                  /estimated_length=100

```

```

misc_feature    43064..51794
                  /note="assembly_fragment"
gap            51795..51894
                  /estimated_length=100
misc_feature    51895..60591
                  /note="assembly_fragment"
gap            60592..61091
                  /estimated_length=100
misc_feature    61092..70105
                  /note="assembly_fragment"
gap            70106..70205
                  /estimated_length=100
misc_feature    70206..77517
                  /note="assembly_fragment"
gap            77518..77617
                  /estimated_length=100
misc_feature    77618..87726
                  /note="assembly_fragment"
gap            87727..87826
                  /estimated_length=100
misc_feature    87827..100723
                  /note="assembly_fragment"
gap            100724..100823
                  /estimated_length=100
misc_feature    100824..115535
                  /note="assembly_fragment"
gap            115536..115635
                  /estimated_length=100
misc_feature    115636..134718

```

Query Match 100.0%; Score 25; DB 14; Length 153267;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATGTGACAC 25
 Db 92634 AAAAAAAAAAGCATGATGTGACAC 92610

```

RESULT 4
AC011466      167272 bp      DNA      linear      PRI 31-JUL-2001
LOCUS        Homo sapiens chromosome 19 clone CTC-453623, complete sequence.
DEFINITION   AC011466
ACCESSION    AC011466.6 GI:15042800
VERSION      HTG.
KEYWORDS     Homo sapiens (human)
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homiidae; Homo
REFERENCE    1 (bases 1 to 167272)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submision
JOURNAL     Unpublished
REFERENCE    2 (bases 1 to 167272)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submision
JOURNAL     Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3 (bases 1 to 167272)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submision
JOURNAL     Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 31, 2001 this sequence version replaced gi:13699565.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center

```

```

www-shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
STS Content:
WI-7187 G06441
SHGC-33567 G28093
WI-14251 G22542
WI-9822 G11762.
Location/Qualifiers
1..167272
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTC-453G23"

ORIGIN

Query Match 100.0%; Score 25; DB 8; Length 167272;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACAC 25
|||||
|||||

DB 133731 AAAAAAAAAAGCATGATTGTGACAC 133755

RESULT 5
AC138524_5
WPCOMMENT
Sequence split into 6 fragments LOCUS AC138524 Accession AC138524
Fragment Name Begin End
AC138524_0 1 110000
AC138524_1 100001 210000
AC138524_2 200001 310000
AC138524_3 300001 410000
AC138524_4 400001 510000
AC138524_5 500001 550277
Continuation 76 of 6) of AC138524 from base 500001 (AC138524 Homo sapiens chromosome 5 c
Query Match 89.6%; Score 22.4; DB 14; Length 50277;
Best Local Similarity 95.8%; Pred. No. 99;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
|||||

DB 2297 AAAAAAAAAAGCATGATTGTGACA 2320

RESULT 6
AC138524_4
WPCOMMENT
Sequence split into 6 fragments LOCUS AC138524 Accession AC138524
Fragment Name Begin End
AC138524_0 1 110000
AC138524_1 100001 210000
AC138524_2 200001 310000
AC138524_3 300001 410000
AC138524_4 400001 510000
AC138524_5 500001 550277
Continuation 75 of 6) of AC138524 from base 400001 (AC138524 Homo sapiens chromosome 5 c
Query Match 89.6%; Score 22.4; DB 14; Length 110000;
Best Local Similarity 95.8%; Pred. No. 77;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
|||||

DB 102297 AAAAAAAAAAGCATGATTGTGACA 102320

RESULT 7
AC093755 146673 bp DNA linear PRI 13-MAY-2005
LOCUS AC093755/c
DEFINITION Homo sapiens BAC clone RP11-55C6 from 4, complete sequence.

```

ACCESSION AC093755 AC025849
VERSION AC093755.2 GI:15638730
KEYWORDS HTG.
SOURCE ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.
1 (bases 1 to 146673)
Maupin, R. and Shah, N.
The sequence of Homo sapiens BAC clone RP11-55C6
Unpublished (2001)
2 (bases 1 to 146673)
Waterston, R.H.
Direct Submission
Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 146673)
Waterston, R.H.
Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 146673)
Waterston, R.
Direct Submission
Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 146673)
Wilson, R.K.
Direct Submission
Submitted (13-MAY-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 18, 2001 this sequence version replaced gi:15529793.

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: H_MH055C06
Drafting Center: WIBR

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30) ; an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPc11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Firengen, E.,
Tateno, M., Catene, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.reagen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-55C6;
actual end is at base position 146673 of RP11-55C6.

FEATURES

The sequence of AC025849 has been incorporated into AC093755.
Location/Qualifiers

source

1.146673
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-55C6"
/clone_1lb="RPCT-11"
misc_feature
81397..81658
/note="CpG island (%GC=62.2, o/e=0.72, #CpGs=21)"
misc_feature
87166..87393
/note="CpG island (%GC=60.5, o/e=0.81, #CpGs=17)"
misc_feature
134005..134230
/note="CpG island (%GC=56.2, o/e=0.89, #CpGs=16)"
ORIGIN

Query Match

Best Local Similarity 95.8%; Score 22.4; DB 8; Length 146673;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
81371 AAAAAAAAAAGCATGATTGTGACA 81348

DB

RESULT 8
AC125167 168056 bp DNA linear HTG 08-MAY-2004
LOCUS
DEFINITION
Mus musculus chromosome 1 clone RP24-445E10, WORKING DRAFT
SEQUENCE, 5 unordered pieces.

AC125167
AC125167
AC125167.3 GI:47084735
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Consensus quality: 166559 bases at least Q30
Consensus quality: 166793 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1.1109: contig of 1109 bp in length
1.1110: 1209: gap of unknown length
1.1210: 43303: contig of 42094 bp in length
43304: 43403: gap of unknown length
43404: 165053: contig of 121650 bp in length
165054: 165153: gap of unknown length
165154: 166579: contig of 1426 bp in length
166580: 166580: gap of unknown length
166580: 168056: contig of 1377 bp in length.

FEATURES

source

1.168056
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/clone="RP24-445E10"
misc_feature
1..1109
/note="assembly_name:Contig10"

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

Consensus quality: 166559 bases at least Q30
Consensus quality: 166793 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1.1109: contig of 1109 bp in length
1.1110: 1209: gap of unknown length
1.1210: 43303: contig of 42094 bp in length
43304: 43403: gap of unknown length
43404: 165053: contig of 121650 bp in length
165054: 165153: gap of unknown length
165154: 166579: contig of 1426 bp in length
166580: 166580: gap of unknown length
166580: 168056: contig of 1377 bp in length.

FEATURES

source

1.168056
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/clone="RP24-445E10"
misc_feature
1..1109
/note="assembly_name:Contig10"

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

gap

misc_feature

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Carliaga, K., Coleman, B., Dietrich, N.L., Engle, J., Graite, S.,
 Guan, X., Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L.,
 Idoi, T.R., Karlins, E., Latic, P., Lee-Lin, S.-Q., Legaepi, R.,
 Maduro, Q.L., Maduro, V.B., Margulies, E.H., Mastello, C., Maskeri, B.,
 Mastralian, S.D., McDowell, J., Paguirigan, C., Pearson, R.,
 Portney, M.E., Prasad, A., Redix-Dugue, N., Schandler, K.,
 Schuster, M.G., Sison, C., Stancirpop, S., Thomas, J.W., Thomas, P.J.,
 Tuchman, J.W., Vogt, J.L., Wetherby, K.D., Wiggins, L., Young, A. and
 Green, E.D.

NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 181432)
 Green, E.D.

Direct Submission
 Submitted (03-OCT-2002) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 3 (bases 1 to 181432)
 Green, E.D.

Direct Submission
 Submitted (01-NOV-2002) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 On Nov 1, 2002 this sequence version replaced gi:23477849.

----- Genome Center -----

Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_zoo@nhgri.nih.gov
 ----- Project Information -----
 Center project name: dbz
 Center clone name: 299K15

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies; the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 179761 bases at least Q40
 Consensus quality: 180204 bases at least Q30
 Consensus quality: 180400 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 180532; sum-of-contigs
 Quality coverage: 11.57x in Q20 bases; agarose-fp
 Quality coverage: 10.76x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 14600: contig of 14600 bp in length
 * 14601: gap of unknown length
 * 14701: contig of 15011 bp in length
 * 14702: gap of unknown length
 * 29712: contig of 3072 bp in length
 * 32883: gap of unknown length
 * 32884: contig of 26433 bp in length
 * 32984: gap of unknown length
 * 59417: gap of unknown length

59517 72053: contig of 12537 bp in length
 * 72054 72153: gap of unknown length
 * 72154 103202: contig of 31049 bp in length
 * 103203 103302: gap of unknown length
 * 103303 145182: contig of 41880 bp in length
 * 145183 145283: gap of unknown length
 * 145284 147255: contig of 1973 bp in length
 * 147256 147355: gap of unknown length
 * 147356 180662: contig of 33307 bp in length
 * 180663 180762: gap of unknown length
 * 180763 181432: contig of 670 bp in length.

FEATURES

source

1..181432
 /organism="Sus scrofa"
 /mol_type="genomic DNA"
 /db_xref="taxon:9823"
 /clone="RP44-299K15"
 /clone_1b="RP44"

misc_feature

1..14600
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"

gap

/estimated_length=unknown
 14701..29711
 /note="assembly_fragment"

gap

29712..29811
 /estimated_length=unknown
 29812..32883
 /note="assembly_fragment"

misc_feature

32884..32983
 /estimated_length=unknown
 32984..59416
 /note="assembly_fragment"

gap

59417..59516
 /estimated_length=unknown
 59517..72053
 /note="assembly_fragment"

misc_feature

72054..72153
 /estimated_length=unknown
 72154..103202
 /note="assembly_fragment"

gap

103203..103302
 /estimated_length=unknown
 103303..145182
 /note="assembly_fragment"

misc_feature

145183..145282
 /estimated_length=unknown
 145283..147255
 /note="assembly_fragment"

gap

147256..147355
 /estimated_length=unknown
 147356..180662
 /note="assembly_fragment"

misc_feature

180663..180762
 /estimated_length=unknown
 180763..181432
 /note="assembly_fragment"

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 181432;
 Best Local Similarity 92.0%; Pred. No. 1.1e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGTGTGACAC 25

Db 6689 AAAAAAAAAAGCATGTGTGACAC 6665

RESULT 10
 AC103436

LOCUS AC103436 21153 bp DNA linear HTG 10-MAY-2003
 DEFINITION Rattus norvegicus clone CH230-158116, *** SEQUENCING IN PROGRESS
 AC103436
 AC103436.6 GI:30521461
 HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 21153)
 Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyadebe, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F.,
 Bielow, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Frazer, C. M., Gabisi, A., Gama, R., Garcia, A., Garner, T., Garrow, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Garcia, A., Gierstra, W., Givara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, J.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, J., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenz, L., Louie, H., Louie, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindrasekaran, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenan, E.,
 Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwokedi, O., Okunolu, G., Olamugbo, A., Pal, S., Parks, K.,
 Pascher, S., Paul, H., Perez, A., Perez, L., Plankoch, C.,
 Plummer, F., Polindexter, A., Popovic, D., Prims, E., Pu, L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, M., Saverly, G., Scherer, S., Scott, G., Shattman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smailes, D.,
 Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Soza, D.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Wilson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weiss, R. A., and Gibbs, R. A.
 Direct Submission
 Title
 Journal
 Unpublished
 2 (bases 1 to 21153)
 Worley, K. C.
 Direct Submission
 Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 21153)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department

COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced gi:25008716.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/ret/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a contig-scaffold). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNL

Center clone name: CH230-158116

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 193830 bases at least Q40

Consensus quality: 196476 bases at least Q30

Consensus quality: 197962 bases at least Q20

Estimated insert size: 210320; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 3 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 208035: contig of 208035 bp in length

* 208036 208135: gap of unknown length

* 208136 209483: contig of 1348 bp in length

* 209484 209583: gap of unknown length

* 209584 211153: contig of 1570 bp in length.

Location/Qualifiers

1. 211153

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-158116"

/note="wgs contig"

misc_feature 125879..126209

misc_feature 40504..41842

misc_feature 125879..126209

misc_feature 40504..41842

misc_feature 125879..126209

misc_feature 40504..41842

misc_feature 125879..126209

misc_feature 40504..41842

misc_feature 125879..126209

misc_feature 40504..41842

ORIGIN

Query Match 87.2% Score 21.8; DB 14; Length 211153;
 Best Local Similarity 92.0%; Pred. No. 1e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
 DB 13268 AAAAAAAAAAGCATGATTGTGACAC 13292

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
RESULT 11 AC098535						
LOCUS	AC098535	300122 bp	DNA	linear	HTG 10-MAY-2003	
DEFINITION	Rattus norvegicus clone CH230-89C2, *** SEQUENCING IN PROGRESS ***.					
ACCESSION	AC098535					
VERSION	AC098535.5 GI:30521670					
KEYWORDS	HTG, HTGS, PHASE1, HTGS_DRAFT, HTGS_ENRICHED.					
SOURCE	Rattus norvegicus					
ORGANISM	Rattus norvegicus (Norway rat)					
REFERENCE						
AUTHORS	Murphy, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalto, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesaar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., De Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Ideblid, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpaty, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowitz, C., Kraft, C. L., Lebow, H., Levin, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maesmar, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangun, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., McNell, A. E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokelam, O., Okwono, G., Olampungason, A., Pal, S., Parks, K., Paternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L. L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Rellay, B., Rellay, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shesman, S., Shen, H., Shetty, J., Shvatsbeyn, A., Slasson, I., Sltter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X. Z., Sorrell, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, T., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, D., Warren, L., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.					
TITLE	Direct Submission					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 300122)					
AUTHORS	Worley, K. C.					
TITLE	Direct Submission					
JOURNAL	Submitted (22-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
REFERENCE	3 (bases 1 to 300122)					

AUTHORS Rat Genome Sequencing Consortium.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On May 10, 2003 this sequence version replaced gi:23268783. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/ratl/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GHIS
 Center clone name: CH230-89C2

----- Summary Statistics
 Assembly program: Atlas 3.0:
 Consensus quality: 247189 bases at least Q40
 Consensus quality: 251780 bases at least Q30
 Consensus quality: 255416 bases at least Q20
 Estimated insert size: 268480; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preeserved.

	1	5293:	contig of 5293 bp in length
*	5294	5393:	gap of unknown length
*	5394	279127:	contig of 273734 bp in length
*	279128	279227:	gap of unknown length
*	279228	289659:	contig of 10432 bp in length
*	289660	289759:	gap of unknown length
*	289760	290895:	contig of 1136 bp in length
*	290896	290995:	gap of unknown length
*	290996	292237:	contig of 1242 bp in length
*	292238	292337:	gap of unknown length
*	292338	293425:	contig of 1088 bp in length
*	293426	293525:	gap of unknown length
*	293526	294886:	contig of 1361 bp in length
*	294887	294986:	gap of unknown length
*	294987	296426:	contig of 1440 bp in length
*	296427	296526:	gap of unknown length
*	296527	300122:	contig of 3596 bp in length.

Location/Qualifiers
 1. 300122
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:101116"
 /clone="CH230-89C2"
 5294. .5393
 gap
 5294. .5393
 /estimated_length=unknown
 5394. .7118
 misc_feature
 /note="wgs contig"
 279128. .279227
 gap


```

/misc_feature
/misc_feature
/misc_feature
misc_feature
misc_feature
gap

ORIGIN
Query Match      85.6%; Score 21.4; DB 14; Length 255295;
Best Local Similarity 95.7%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAAAAAAAAAGCATGATTGTGACA 24
Db      134035 AAAAAAAAAAGCATGATTGTGACA 134057

RESULT 13
LOCUS      CO868604      32404 bp      DNA      linear      PAT 13-SEP-2004
DEFINITION Sequence 16 from Patent WO2004074321.
ACCESSION  CO868604
VERSION     CO868604.1 GI:51998601
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE   1 Morris, D.W. and Malandro, M.S.
            Novel therapeutic gpcr targets in cancer
            Patent: WO 2004074321-A 16 02-SEP-2004;
            Sagres Discovery, Inc. (US)
FEATURES
            source
            1. .32404
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"

ORIGIN
Query Match      84.0%; Score 21; DB 6; Length 32404;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTG 21
Db      302 AAAAAAAAAAGCATGATTGTG 322

RESULT 14
LOCUS      AX695785      32404 bp      DNA      linear      PAT 31-MAR-2003
DEFINITION Sequence 1412 from Patent WO03008583.
ACCESSION  AX695785
VERSION     AX695785.1 GI:29418939
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
REFERENCE   1

```

```

AUTHORS      Morris, D.W. and Engelhard, E.K.
TITLE        Novel compositions and methods for cancer
            Patent: WO 03008583-A 1412 30-JAN-2003;
            Sagres Discovery (US)
FEATURES
            source
            1. .32404
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"

ORIGIN
Query Match      84.0%; Score 21; DB 6; Length 32404;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTG 21
Db      302 AAAAAAAAAAGCATGATTGTG 322

RESULT 15
LOCUS      BV018408      819 bp      DNA      linear      STS 30-MAY-2003
DEFINITION S212P6120F68.T0 CZECHII/El Mus musculus STS genomic, sequence
            tagged site.
ACCESSION  BV018408
VERSION     BV018408.1 GI:31102303
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 819)
REFERENCE   1 Wender, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,
            Lander, E.S., Lindblad-Toh, K. and Daly, M.J.
            The mosaic structure of variation in the laboratory mouse genome
            Nature 420 (6915), 574-578 (2002)
            12466852
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kerstin@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 819
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 12951/Syrm, CH/Ref, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSVC3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
            Location/Qualifiers
            1. .819
               /organism="Mus musculus"
               /mol_type="genomic DNA"
               /strain="CZECHII/El"
               /db_xref="taxon:10090"
               /map="10 22-736 80561980-80561266"
               /clone_lib="CZECHII/El"
               <1. .>819

ORIGIN
Query Match      83.2%; Score 20.8; DB 10; Length 819;
Best Local Similarity 91.7%; Pred. No. 1.5e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
 |||||
 781 AAAAAAAAAAGCATGATTGTGACA 804

RESULT 16
 AX506824/c
 LOCUS AX506824 2013 bp DNA linear PAT 27-SEP-2002
 DEFINITION Sequence 1519 from Patent WO0216655.
 ACCESSION AX506824
 VERSION AX506824.1 GI:23388061
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
 TITLE Stress-regulated genes of plants, transgenic plants containing
 same, and methods of use
 JOURNAL Patent: WO 0216655-A 1519 28-FEB-2002;
 The Scripps Research Institute (US) ; Syngenta Participations AG
 (CH)

FEATURES
 source Location/Qualifiers
 1..2013 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 6; Length 2013;
 Best Local Similarity 91.7%; Pred. No. 1.1e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
 |||||
 1329 AAAAAAAAAAGCATGATTGTGACA 1306

RESULT 17
 BX511033/c
 LOCUS BX511033 51202 bp DNA linear VRT 03-APR-2004
 DEFINITION zebrafish DNA sequence from clone CH211-116M9 in linkage group 5,
 complete sequence.
 ACCESSION BX511033
 VERSION BX511033.5 GI:46200422
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 51202)
 Hunter, G.
 Direct Submission
 Submitted (03-APR-2004) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Apr 5, 2004 this sequence version replaced g1:45597966.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk

COMMENT

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information
 on the WormPeP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep/Clone-derived
 zebrafish pUC subclones occasionally display inconsistency over the
 length of mononucleotide A/T runs and conserved TA repeats. Where
 this is found the longest good quality representation will be
 submitted.
 Repeat names beginning 'Dr' were identified by the Recon repeat
 discovery system (Zhirony Bao and Sean Eddy, submitted), and those
 beginning 'drr' were identified by Rick Waterman (Stephen Johnson
 lab, WashU). For further information see
 http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-116M9
 is from a CHORI-211 BAC library
 VECTOR: pTRBAC2.1.

FEATURES
 source Location/Qualifiers
 1..51202 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-116M9"
 /clone_1ib="CHORI-211"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 5; Length 51202;
 Best Local Similarity 91.7%; Pred. No. 3.9e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGATTGTGACAC 25
 |||||
 31469 AAAAAAAAAAGCATGATTGTGACAC 31446

RESULT 18
 CT009634/c
 LOCUS CT009634 96729 bp DNA linear HTG 12-AUG-2005
 DEFINITION Danio rerio chromosome 5 clone CH73-36J8, WORKING DRAFT SEQUENCE,
 6 unordered pieces.
 ACCESSION CT009634
 VERSION CT009634.2 GI:72533916
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 96729)
 McLaren, S.
 Direct Submission
 Submitted (11-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests:
 http://www.sanger.ac.uk/Projects/D_rerio/fage.shtml#data1gthc
 On Aug 12, 2005 this sequence version replaced g1:56687942.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: ZH36J8
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads

Consensus quality: 94801 bases at least Q40
Consensus quality: 95167 bases at least Q30
Consensus quality: 95524 bases at least Q20
Insert size: 96229; sum-of-contigs
Insert size: 104769; 4.4% error; agarose-fp
Quality coverage: 8.92% in Q20 bases; sum-of-contigs Quality
coverage: 8.29% in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 3956: contig of 3956 bp in length
* 3957 4056: gap of 100 bp
* 4057 21454: contig of 17398 bp in length
* 21455 21554: gap of 100 bp
* 21555 29844: contig of 8290 bp in length
* 29845 32939: contig of 2995 bp in length
* 32940 33039: gap of 100 bp
* 33040 44060: contig of 11021 bp in length
* 44061 44160: gap of 100 bp
* 44161 96729: contig of 52569 bp in length.

FEATURES
source
1. 96729
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="5"
/clone="CH73-366J8"
/clone_1id="CHOR1-73"
1. 3956
/note="assembly: fragment:00020
fragment_chain:1
clone_end:SP6
vector_side:left"

misc_feature

1. 3956
/note="assembly: fragment:00020
fragment_chain:1
clone_end:SP6
vector_side:left"

misc_feature

4057. 21454
/note="assembly: fragment:00267
fragment_chain:1"

misc_feature

21555. 29844
/note="assembly: fragment:00075
fragment_chain:1"

misc_feature

29945. 32939
/note="assembly: fragment:00007
fragment_chain:1"

misc_feature

33040. 44060
/note="assembly: fragment:00151
fragment_chain:1"

misc_feature

44161. 96729
/note="assembly: fragment:00550
fragment_chain:1
clone_end:T7
vector_side:right"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 96729;
Best Local Similarity 91.7%; Pred. No. 3.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
|||||

Db 39600 AAAAAAAAAAGCATGATTGTGACA 39577
|||||
|||||

RESULT 19 F19P19 107527 bp DNA linear PLN 11-MAR-1999
LOCUS F19P19
DEFINITION Sequence of BAC F19P19 from Arabidopsis thaliana chromosome 1,
complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC000104
AC000104.1 GI:2341023
HTG.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS

Oborne, B.I., Vysotskaia, V.S., Schwartz, J.R., Tortum, M., Yu, G.,
Buehler, E., Conway, A.B., Conway, A.R., Dewar, K., Peng, J., Kim, C.,
Kurtz, D., Li, Y., Shim, P., Sun, H., Davis, R.W., Becker, J.R.,
Fiederspigel, N.A. and Theologis, A.
The sequence of BAC F19P19 from Arabidopsis thaliana chromosome 1
Unpublished (1999)

REFERENCE
AUTHORS

2 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (06-JAN-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

REFERENCE
AUTHORS

3 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (30-JAN-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

REFERENCE
AUTHORS

4 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (21-MAR-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

REFERENCE
AUTHORS

5 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (16-APR-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

REFERENCE
AUTHORS

6 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (22-AUG-1997) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA

REFERENCE
AUTHORS

7 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (26-AUG-1997) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA

REFERENCE
AUTHORS

8 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (16-SEP-1997) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA

REFERENCE
AUTHORS

9 (bases 1 to 107527)
Theologis, A.

TITLE
JOURNAL

Direct Submission
Submitted (11-MAR-1999) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA

REFERENCE
AUTHORS

On Aug 22, 1997 this sequence version replaced gi:1808697.
The sequence of BAC F19P19 from Arabidopsis thaliana chromosome 1.

COMMENT

FEATURES
source

1. 107527
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F19P19"
/ecotype="Columbia"
/gene="F19P19.1"
complement(208..599,932..1352)
/gene="F19P19.1"
/codon_start=1
/evidence=not experimental
/product="F19P19.1"
/protein_id="AAB70422.1"

/db_xref="GI:2341024"
/translation="MGLSISFLSIIMMCLLPEDLVNVKSAATTEYTLIKGCARQO
FSDPSGLVQALSMFSGSLVQSSTKTRFKYTKTGSTTTTGLPQCRGLDSHNDGFC
VRLPVLSDKLCGKTIASRVQSGCLLVEGSPGSISEMELFKTCGKNLADGFC
ERRDAFGVMONGVSGHGFATTVYSSVVLQCCEDVDGDTDCGCVKNALEKQVVE
CGSISGQIYLHCFLIASYYPNGVPRRSSSSSSSSSSSSSSSSSSSSSDPSTSTG"
complement(2612..6455)
/gene="F19P19.2"
/complement(join(2612..2739,2831..2901,2964..3087,
3227..3311,3424..3531,3581..3700,4248..4307,4403..4498,
4757..4873,5037..5158,5205..5421,5538..5624,5791..5826,
5895..6000,6046..6212,6267..6352,6443..6455))
/gene="F19P19.2"
/codon_start=1
/evidence=not experimental
/protein_id="AAB70423.1"
/db_xref="GI:2341025"
/translation="MNCASGEVPEEPYVSKSGULYEKRLIQTHSLFSLHLSTCC
RFLNDFVVIITRLILVPCRIWGNARLIVSHLLMTLFPSSKLRNLFKPSISGTL
QIVKPPPTASTIPGLLGTFOQDHIELKMNLIWENDSLMNSFALBQOLHARQELS
HALYVIDGYTFPLQHDACRYIARLKKERDESQDLAEAROLPAEAVATSNAL
SNGKRMALARLLNLVSMLCGIDBQGPNAKMRIGISAELITELDCNAALSOORAK
ROIPKTLASVDALKEFTOLSHPIKTKNGKGFMSMDILSKOVIATGGIDTTAVLFD
PSGOLSTVIGSKKVTYSIKVQDPTLVLTASDKVTASDENDVATTAAPRQGLIL
GRTAOSITIKINDVKSQFVLTSLSLSPDRQMSLVDTMEKSPVHSLTTAALD
GRWLMDRLKKNRFTFPDANSGEVYPTQCEWELIDGAHFDNEILKOSCANSLNL
SPMOGVFOAASVKAMNPKITLPDLSGTGATSVKFGIDSKYIAVGSMDRLRIFG
LPDDDTSDSAQOS"
9724..11278
/gene="F19P19.3"
/join(9724..9745,9861..9958,10023..10715,10839..11069,
11123..11278)
/gene="F19P19.3"
/codon_start=1
/evidence=not experimental
/product="F19P19.3"
/protein_id="AAB70424.1"
/db_xref="GI:2341026"
/translation="METLPPYRLFVIVSCFTRKSHQALAFYASLQMSINKOLDDEIT
/LTAQIIPFCDSOLFQETPNOTSEVTSASNGCYENNNTNFPDKSGSNODIED
NNDNADLSIIPSDODFNDITASIDPSSIOBPADLOBOFPDFTGLDHPPTLY
SSSGDULPPLISVPEEDCLSSVPSNLGINSIPSSSCFLNTGTPPTMYTGMNM
TGIGSGFTSGNTHLGSDFKPSHDQMEIQADNGDLCPDPIKFTFPGDHIQGLGV
EKQNHVNAQVPLPQLGTEITGLDPSFNVGKLSAQOREKILHRVKKENENFSKIL
KVFVLNHNRIKPVACRKTLDASRPRVGRFAKNDPGRQACSHHEDDDVDR
LILFDYN"
13879..15269
/gene="F19P19.4"
/join(13879..14922,15023..15175,15228..15269)
/gene="F19P19.4"
/note="EST gb|T21788 comes from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AAB70425.1"
/db_xref="GI:2341027"
/translation="MEGQTHDAQNNVPEHDNPLMANKPSSNIFSKDAQCSVLGNG
KMQLIVRSPKVSCEASTCPSHILTVGNRYVAESKMKKVIISFGSKTSQGLDFT
PMAMGTPEHVFLENOQDYVASLNVSDONSMTIHLVKGESASASSLVRINVT
LFEKVLREBFVLFSNGENLKIPTRKARGSKYVHVRERRARLSRISFIDL
GSMDEFOAONVDCILKNKLPNTLFTLAVVVOETTEDRIGMGKLFKLRSRPOGNS
DASERTSTSSISMNVVIPSCHIGSPEDPSSILBWKQACNDCGNDLCCSLTLK
GQPRKQYFELFTBGSKHETTGKTIIVNSGGLVQLVQFEKLTLSQSFALALAFHSEK
LPLHLKLEKCOILERRSF"
complement(15947..17024)
/gene="F19P19.5"
/complement(join(15947..16071,16676..17024))
/gene="F19P19.5"
/note="Strong similarity to 60S ribosomal protein L17
(gb|X01694). EST gb|AA042332 comes from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AAB70426.1"
/db_xref="GI:2341028"

/translation="MCLFSFSLVLIIVVDLDFRLRGQGTSGNKRMSIQLPVAATVN
CADNTGAKNLYIISVKGIGKRLNRLPSACVGMVAVYKKGPRDKVLPVIVRQR
KPMRRKDVFWYFEDNAGIVIPKEMGSAITGPIGKXCADLMPRIASANAIV"
18613..22027
/gene="F19P19.6"
/join(18613..19224,19307..19974,20046..20718,20796..20982,
21060..22027)
/gene="F19P19.6"
/note="EST gb|ATTS5672 comes from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AAB70427.1"
/db_xref="GI:1903347"
/translation="MVRHSRRSPFDWYGFVCPDPTDLMPKGLDGLDRDEIREYAY
EIFFAACSSPGFGRNALTFYSKNAGHQDGIQGGGSGSSGSGFGSGIKRKL
TTPSRVRLGLKMLKSPSRMSVTGVVAVAPASPGNNGISGSGSPFAG
PFTVPARPRPRTSARTIMROOMKYTBSQDRLRTIMTLYVGGRAPRTILPEL
LHVXPSRFGVDHAEQIMORQKVLBAGLIIHPSTPLEKTNPFMRRLREIROSRTK
ADTVKSDIEMLEKNTLVASLSWRNATPTTDICHAADGIPNIIHLVADLQSIPTDRL
ETLVDEIDELMLKKTWIMIGITRAIHNLCTFWVLPHQYVTSQMEPDLGASHAM
LAEVNDKASDREALVYLLTSTLASQGWTEKRLSHDYFORGNVGLINLLPLA
LSSKILGSDVTISQNGLEKQDVLVDSGDRVDYIRASIKNASVYIENMKAEIE
ETBEGEREATWMLLAKETEDLALRSCFSPILKRMHVAAGVASVLHOCYGSIL
MOYLRGRTITTEYEVQTAGKLEKVLQVYVAEASDECDGKGLVPRVAYEVDI
IURLROMITEKLOVQECTSRAKATNPNPKSKSPYIQSGRGLMKLANDAIEEPE
IPIGITEDLVHDLAEGLELFGQVTFPVASCOSKOSYITLPLTRCNDSKPVKMK
KATPVAAGEELQWGEAPGNGHPPRSRGTQRYIRLNTLHFLBSQHSJNKSLSL
NPRVLPATRKCRERTKSSYEPFOAGIESACQHVSEVAYRLIFLDSYVYFESLY
PGDVANGRIKPALRIKONLITMTAQLIDKALAMKEKMAKSFVYLVLAAQGRS
VRCRTHDLIEEDFESLKRYVCTGCEGLIPREVLREAEVGVQIMQGPTEQMED
FSIVTCSGSMGLVGTGQLPMPPTTGKRRNSDPNTILRVLCYRDDRVANQPLKKSFO
LGKR"
22508..22885
/gene="F19P19.7"
/join(22508..22520,22653..22885)
/gene="F19P19.7"
/note="Strong similarity to 50S ribosomal protein L28
(gb|X68078)."
/codon_start=1
/evidence=not experimental
/protein_id="AAB70428.1"
/db_xref="GI:1903348"
/translation="MLPSRICPPTGKANKRANKVSFSSHHKTKKQLQVNLQYKRWME
AKRRVILRLSTKALTIENGLBAVSKAGIDLRKK"
complement(23351..24198)
/gene="F19P19.8"
/complement(join(23351..23802,24144..24198))
/gene="F19P19.8"
/codon_start=1
/evidence=not experimental
/product="F19P19.8"

Query Match 83.2%; Score 20.8; DB 15; Length 107527;
Best Local Similarity 91.7%; Pred No. 36+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGCATGATGTGACA 24
|||||
Db 32791 AAAAAAAAAAAGCATGATGTGACA 32814
|||||

RESULT 20
AC094469 0
WPCOMMENT
Sequence split into 4 fragments
Fragment Name Begin End
AC094469_0 1 110000
AC094469_1 100001 210000
AC094469_2 200001 310000
AC094469_3 300001 366202
LOCUS AC094469 366202 bp DNA linear HTG 22-SEP-2002
DEFINITION Rattus norvegicus clone CH230-4B3, *** SEQUENCING IN PROGRESS ***,
7 unordered pieces.

ACCESSION AC094469
 VERSION AC094469.10 GI:23268176
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurgnathi; Muroidae; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 366202)
 1 Muny, D., Marie, M., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalbechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, D., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Centes, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Delgado, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Diaper, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Joliver, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Joliver, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koval, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louisgeed, H., Lozada, R., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Manthey, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Narkavicius, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackemeh, O., Okwomu, G., Olarnpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S., J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waidron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhasern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 366202)
 2 Worley, K.C.
 Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 366202)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Sep 22, 2002 this sequence version replaced gi:22772348.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc_help@bcm.tmc.edu

Project Information

Center project name: GASD

Center clone name: CH230-4B3

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 241208 bases at least Q40

Consensus quality: 243075 bases at least Q30

Consensus quality: 244806 bases at least Q20

Estimated insert size: 270965; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 340986: contig of 340986 bp in length

* 340987 341086: gap of unknown length

* 341087 344539: contig of 3453 bp in length

* 344540 344639: gap of unknown length

* 344640 349127: contig of 4488 bp in length

* 349128 349227: gap of unknown length

* 349228 361781: contig of 12554 bp in length

* 361782 361881: gap of unknown length

* 361882 363463: contig of 1582 bp in length

* 363464 363563: gap of unknown length

* 363564 364872: contig of 1309 bp in length

* 364873 364972: gap of unknown length

* 364973 366202: contig of 1230 bp in length.

* Location/Qualifiers

FEATURES

source

1. 366202

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-4B3"

1. 1064

/note="wgs_end extension

clone_end:Sp6"

4222. 5040

/note="clone boundary

clone_end:Sp6"

site:ECORI

end_sequence:BH303958"

339617. 340986

/note="wgs contig"

340987. 341086

/estimated_length=unknown

344540. 344639

/estimated_length=unknown

344640. 346412

/note="wgs contig"

349128. 349227

/estimated_length=unknown

349228. 350614

/note="wgs contig"

misc_feature

```

gap          361782..361881
gap          /estimated_length=unknown
gap          363464..363563
gap          /estimated_length=unknown
gap          364873..364972
gap          /estimated_length=unknown

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 110000;
Best Local Similarity 91.7%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTGACA 24
Db      77357 AAAAAAAAAACCTGATTGTGACA 77380

RESULT 21
BX908758.1
WPCOMMENT
Sequence split into 6 fragments LOCUS BX908758 Accession BX908758
Fragment Name      Begin      End
BX908758-0         1      110000
BX908758-1         100001   210000
BX908758-2         200001   310000
BX908758-3         300001   410000
BX908758-4         400001   510000
BX908758-5         500001   579882
Continuation (2 of 6) of BX908758 from Base 100001 (BX908758 Danio rerio clone DKEX-3262)

Query Match      83.2%; Score 20.8; DB 14; Length 110000;
Best Local Similarity 91.7%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTGACA 24
Db      33363 AAAAAAAAAAGCATGAAAGTGACA 33386

RESULT 22
AC149546
LOCUS      AC149546      124127 bp      DNA      linear      PLN 07-APR-2005
DEFINITION Medicago truncatula clone mch2-70013, complete sequence.
ACCESSION  AC149546
VERSION     AC149546.14 GI:62238013
KEYWORDS   HTG.
SOURCE      Medicago truncatula (barrel medic)
ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
            Medicago.
REFERENCE   1 (bases 1 to 124127)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
TITLES    Medicago truncatula BAC Clone mch2-70013
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 124127)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
TITLES    Direct Submision
JOURNAL   Submitted (05-JUN-2004) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE   3 (bases 1 to 124127)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
TITLES    Direct Submision
JOURNAL   Submitted (06-APR-2005) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE   4 (bases 1 to 124127)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.

```

```

TITLES    and Roe,B.A.
JOURNAL   Direct Submision
            Submitted (07-APR-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT   On Apr 6, 2005 this sequence version replaced gi:59676736.
            ----- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            -----
FEATURES             Location/Qualifiers
     source            1..124127
                        /organism="Medicago truncatula"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:3880"
                        /clone="mch2-70013"
                        /clone_lib="Medicago truncatula BAC library H2"

ORIGIN
Query Match      83.2%; Score 20.8; DB 15; Length 124127;
Best Local Similarity 91.7%; Pred. No. 2.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTGACA 24
Db      112876 AAAAAAAAAATCATGATTGTGACA 112899

RESULT 23
AC109193/C
LOCUS      AC109193      144093 bp      DNA      linear      ROD 31-JUL-2004
DEFINITION Mus musculus chromosome 5, clone RP24-352122, complete sequence.
ACCESSION  AC109193
VERSION     AC109193.15 GI:50872577
KEYWORDS   HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 144093)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
JOURNAL   Unpublished
TITLES    Mus musculus chromosome 5, clone RP24-352122
REFERENCE   2 (bases 1 to 144093)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B.,
            Brown,A., Camarata,J., Campiolo,A., Chang,J., Chazaro,B.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,C., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Poljara V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rossetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talams,J., Testaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Triggilo,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLES    Direct Submision
JOURNAL   Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 144093)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

```

TITLE	Anderson, M., Anderson, S., Arachchi, H.M., Barua, N., Baetien, V., Bloom, T., Boguslavsky, L., Bourkgatler, B., Camarata, J., Chang, J., Chepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Derrillano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lander, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneses, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Notbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Plierre, N., Rachpuka, A., Ransamy, U., Raymond, C., Retta, R., Riese, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Roman, J., Schauer, S., Schupack, R., Stojanovic, N., Stubbs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliou, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
JOURNAL	Submitted (09-JUN-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	4 (bases 1 to 144093)
AUTHORS	Birtten, B., Nisbaum, C., Lander, E., Aboueljelil, A., Allen, N.,

TITLE	JOURNAL	REFERENCE	AUTHORS
Direct Submission	Submitted (09-JUN-2004)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	4 (bases 1 to 144093)
			Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Chepel, Y., Collymore, A., Cook, P., Corum, B., DeArrellano, K., Diaz, J.S., Dode, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafer, N., Hagopian, D., Hagos, B., Hall, U., Horton, L., Hulme, W., Iliev, N., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stengs-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission	Submitted (31-JUN-2004)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	On Jul 31, 2004 this sequence version replaced gi:48475321.
			All repeats were identified using RepeatMasker:

```

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (31-JUL-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 31, 2004 this sequence version replaced gi:48475321.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center -----
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@road.mit.edu
----- Project Information -----
Center project name: L20189
Center clone name: 352_I_22
-----
FEATURES
source
Location/Qualifiers
1. 144093
/oranism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="5"
/map="5"
/clone="RP24-352I22"
/clone_lib="RPCT-24 Male Mouse BAC"
1..4
/note="clone_boundary"
/clone_end:SP6

```

repeat_region	site:Wb01"	complement(769. ,.903)
repeat_region	/rpt_family="Lx8"	complement(904. ,1249)
repeat_region	/rpt_family="ORR1A3"	complement(1250. ,1326)
repeat_region	/rpt_family="Lx8"	1756. ,1800
repeat_region	/rpt_family="CAAAC"	complement(1909. ,2041)
repeat_region	/rpt_family="B2_Mm2"	complement(2791. ,2862)
repeat_region	/rpt_family="ID_B1"	complement(2833. ,2893)
repeat_region	/rpt_family="B1F"	complement(2861. ,2917)
repeat_region	/rpt_family="B1_Mm1"	complement(2924. ,3048)
repeat_region	/rpt_family="ID_B1"	complement(3361. ,3426)
repeat_region	/rpt_family="M1R"	complement(3614. ,3811)
repeat_region	/rpt_family="Lx"	3665. ,3900
repeat_region	/rpt_family="AGGCGG"	complement(3906. ,3972)
repeat_region	/rpt_family="PB1D9"	3973. ,4143
repeat_region	/rpt_family="TTCTC"	4145. ,4274
repeat_region	/rpt_family="CT-rich"	complement(4286. ,4350)
repeat_region	/rpt_family="cPNA-ALA-GCA"	complement(4475. ,4583)
repeat_region	/rpt_family="Lx9"	complement(4615. ,5053)
repeat_region	/rpt_family="Lx9"	5086. ,5166
repeat_region	/rpt_family="TATG"	5417. ,5541
repeat_region	/rpt_family="CA"	6235. ,6285
repeat_region	/rpt_family="CA"	6674. ,7048
repeat_region	/rpt_family="CATATTA"	7051. ,7281
repeat_region	/rpt_family="CA"	complement(7330. ,7702)
repeat_region	/rpt_family="ORR1D"	8056. ,8106
repeat_region	/rpt_family="(TG)"	complement(8115. ,8231)
repeat_region	/rpt_family="RSINB1"	9190. ,9341
repeat_region	/rpt_family="ID_B1"	9285. ,9380
repeat_region	/rpt_family="PB1D10"	9650. ,9693
repeat_region	/rpt_family="(TG)"	complement(9712. ,9846)
repeat_region	/rpt_family="MTE"	9873. ,9957
repeat_region	/rpt_family="MERSA"	10135. ,10302
repeat_region	/rpt_family="(CA)"	10321. ,10472
repeat_region	/rpt_family="(CA)"	10473. ,10531
repeat_region	/rpt_family="(TC)"	10531. ,10566
repeat_region	/rpt_family="(CA)"	complement(10998. ,11108)
repeat_region	/rpt_family="MERS2"	

```

repeat_region      11421..11472
                    /rpt_family="(TG)n"
repeat_region      complement(11475..11692)
                    /rpt_family="B4A"
repeat_region      12353..12452
                    /rpt_family="MIR"
repeat_region      complement(12553..12739)
                    /rpt_family="LIM3"
repeat_region      12849..13246
                    /rpt_family="RMR19B"
repeat_region      13273..13411
                    /rpt_family="RMR19B"

Query Match      83.2%; Score 20.8; DB 9; Length 144093;
Best Local Similarity 91.7%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAAGCATGATTGTGACA 24
Db      30532 AAAAAAAAAAGCATGATTGTGACA 30509

RESULT 24
LOCUS      CR354586      144375 bp      DNA      linear      HTG 21-MAR-2005
DEFINITION      Danio rerio clone CH211-142B20, WORKING DRAFT SEQUENCE, 10
unordered pieces.
ACCESSION      CR354586
VERSION      CR354586.5 GI:61673514
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 144375)
McLaren, S.
Direct Submission
Submitted (18-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 2005 this sequence version replaced gi:45598153.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZC142B20
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 140573 bases at least Q40
Consensus quality: 141166 bases at least Q30
Consensus quality: 141642 bases at least Q20
Insert size: 143475; sum-of-contigs
Insert size: 149956; 0.9% error; agarose-fp
Quality coverage: 8.43x in Q20 bases; sum-of-contigs Quality
coverage: 8.14x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 31090: contig of 31090 bp in length
* 31091 31190: gap of 100 bp
* 31191 34144: contig of 2954 bp in length
* 34145 34244: gap of 100 bp
* 34245 64098: contig of 29854 bp in length

```

```

* 64099 64198: gap of 100 bp
* 64199 71649: contig of 7451 bp in length
* 71650 71749: gap of 100 bp
* 71750 82168: contig of 10419 bp in length
* 82169 82268: gap of 100 bp
* 82269 91679: contig of 9411 bp in length
* 91680 91779: gap of 100 bp
* 91780 113436: contig of 21657 bp in length
* 113437 113536: gap of 100 bp
* 113537 136646: contig of 23110 bp in length
* 136647 136746: gap of 100 bp
* 136747 139095: contig of 2349 bp in length
* 139096 139195: gap of 100 bp
* 139196 144375: contig of 5180 bp in length.
Location/Qualifiers
1..144375
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-142B20"
/clone_1fb="CHORI-211"
1..31090
/note="assembly fragment:01825
fragment chain:1
clone_end:T7
vector_side:left"
31191..34144
/note="assembly fragment:00039
fragment chain:1"
34245..64098
/note="assembly fragment:01338
fragment chain:1"
64199..71649
/note="assembly fragment:00091
fragment chain:1"
71750..82168
/note="assembly fragment:00227
fragment chain:1"
82269..91679
/note="assembly fragment:00337
fragment chain:1"
91780..113436
/note="assembly fragment:00489
fragment chain:1"
113537..136646
/note="assembly fragment:00909
fragment chain:1"
136747..139095
/note="assembly fragment:00006"
139196..144375
/note="assembly fragment:00150
clone_end:Sp6
vector_side:right"

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 144375;
Best Local Similarity 91.7%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAAAAAAAAAGCATGATTGTGACA 24
Db      90838 AAAAAAAAAAGCATGATTGTGACA 90861

RESULT 25
LOCUS      AC138453      144639 bp      DNA      linear      PLN 15-AUG-2003
DEFINITION      Medicago truncatula clone mch2-20m14, complete sequence.
ACCESSION      AC138453
VERSION      AC138453.8 GI:33589888
KEYWORDS      HTG.
SOURCE      Medicago truncatula (barrel medic)
ORGANISM      Medicago truncatula

```

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosoid I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE
AUTHORS 1 (bases 1 to 144639)
TITLE Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
JOURNAL Cook,D., Kim,D. and Roe,B.A.
REFERENCE
AUTHORS Medicago truncatula BAC Clone mth2-20m14
TITLE Unpublished
JOURNAL 2 (bases 1 to 144639)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
TITLE Cook,D., Kim,D. and Roe,B.A.
JOURNAL Direct Submission
REFERENCE
AUTHORS Submitted (03-JAN-2003) Department Of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 3 (bases 1 to 144639)
TITLE Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
JOURNAL Cook,D., Kim,D. and Roe,B.A.
REFERENCE
AUTHORS 4 (bases 1 to 144639)
TITLE Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
JOURNAL Cook,D., Kim,D. and Roe,B.A.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (12-AUG-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 5 (bases 1 to 144639)
TITLE Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
JOURNAL Cook,D., Kim,D. and Roe,B.A.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (15-AUG-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Aug 12, 2003 this sequence version replaced gi:33354337.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
source Location/Qualifiers
1..144639
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-20m14"
/clone_lib="Medicago truncatula BAC library H2"
ORIGIN
Query Match 83.2%; Score 20.8; DB 15; Length 144639;
Best Local Similarity: 91.7%; Pred. No. 2,8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 51290 AAAAAAAAAATCATGATTGTACA 51313
RESULT 26
LOCUS BX957226 149414 bp DNA linear VRT 01-SEP-2004
DEFINITION zebrafish DNA sequence from clone CH211-110P13 in linkage group 5,
complete sequence.
BX957226
BX957226.10 GI:51127576
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.
REFERENCE
AUTHORS 1 (bases 1 to 149414)
TITLE Dunn,M.
JOURNAL Direct Submission
REFERENCE
AUTHORS Submitted (01-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 11, 2004 this sequence version replaced gi:51035920.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WormPep; Information
on the WormPep database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormPep
Zebrafish PUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhirong Bao and Sean Eddy, submitted), and those
beginning 'drr' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
CH211-110P13 is from a CHOR1-211 BAC library
VECTOR: pTARBA2.1.

FEATURES
source Location/Qualifiers
1..149414
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-110P13"
/clone_lib="CHOR1-211"
ORIGIN
Query Match 83.2%; Score 20.8; DB 5; Length 149414;
Best Local Similarity: 91.7%; Pred. No. 2,7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 92733 AAAAAAAAAAGCATGATTTTTACA 92756
RESULT 27
LOCUS AL606970/c 161863 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-503C24 on chromosome 6 contains
the 3' end of a novel gene, up to five novel genes, (including
FLJ1232 and ppi3671), a meningioma expressed antigen 6
(coiled-coil proline-rich) (MGEA6) pseudogene and three Cpg
islands, complete sequence.
AL606970
AL606970.12 GI:17065766
HTG: Cpg island; FLJ1232; MGEA6; ppi3671.

```
SOURCE      Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominoidea; Homo.
REFERENCE    1 (bases 1 to 161863)
AUTHORS      Tracey, A.
TITLES       Direct Submision
              Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
JOURNAL      Clome requests: clome@sanger.ac.uk
              On Nov 25, 2001 this sequence version replaced gi:17017816.
              The following abbreviations are used to associate primary accession
              numbers given in the feature table with their source databases:
              Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
              on the WORMPEP database can be found at
              http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
              was generated from part of bacterial clone contigs of human
              chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
              Group. Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr6
              RP11-503C24 is from the library RPCI-11.2 constructed by the group
              of Pieter de Jong. For further details see
              http://www.chori.org/bacpac/home.htm
              VECTOR: pBACE3.6
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: vegas@sanger.ac.uk
              -----
              This sequence was finished as follows unless otherwise noted: all
              regions were either double-stranded or sequenced with an alternate
              chemistry or covered by high quality data (i.e., phred quality >=
              30); an attempt was made to resolve all sequencing problems, such
              as compressions and repeats; all regions were covered by at least
              one subclone; and the assembly was confirmed by restriction digest,
              except on the rare occasion of the clone being a YAC.
              Location/Qualifiers
                source          1..161863
                                /organism="Homo sapiens"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:9606"
                                /chromosome="6"
                                /clone_lib="RP11-503C24"
                                /clone_1ib="RPCI-11.2"
                                1
                                /note="Clone left end: RP11-503C24"
                                /join(complement(AL611929.2:9197..9527),2500..2617)
                                /locus_tag="RP11-503C24.1-001"
                                /join(complement(AL611929.2:9197..9527),2500..2617)
                                /locus_tag="RP11-503C24.1-001"
                                /note="match: ESTs: A149978"
                                complement(join(23395..23640,24425..24633))
                                /locus_tag="RP11-503C24.2-001"
                                complement(join(23395..23640,24425..24633))
                                /locus_tag="RP11-503C24.2-001"
                                /locus_tag="RP11-503C24.2-001"
                                /note="match: ESTs: A1825113 AW295767"
                                join(27848..27970,37138..37225,38570..38831)
                                /locus_tag="RP11-503C24.3-001"
                                join(27848..27970,37138..37225,38570..38831)
                                /locus_tag="RP11-503C24.3-001"
                                /note="match: ESTs: A1829659"
                                33700..33768
                                /locus_tag="RP11-503C24.3-001"
                                /note="Single clone region. Sequence from reads from a
                                short insert library derived from a single pUC clone.
                                Restriction digest data confirm the assembly."
                                34333..34503
                                /locus_tag="RP11-503C24.3-001"
                                /note="Single clone region. Sequence from reads from a
                                short insert library derived from a single pUC clone.
                                Restriction digest data confirm the assembly."
              misc_feature
              34695
                /locus_tag="RP11-503C24.3-001"
                /note="Random repeat. Forced join. Gap size estimated to
                be approximately 500bp by restriction digest data."
                34825..35086
                /locus_tag="RP11-503C24.3-001"
                /note="Single clone region. Sequence from reads from a
                short insert library derived from a single pUC clone.
                Restriction digest data confirm the assembly."
                join(41054..41191,56984..57492,57781..58006,58784..59486,
                60028..60697)
                /locus_tag="RP11-503C24.4-001"
                join(41054..41191,56984..57492,57781..58006,58784..59486,
                60028..60697)
                /locus_tag="RP11-503C24.4-001"
                /locus_tag="RP11-503C24.5-001"
                /note="match: CDNA: AK055794"
                63376
                /note="Random repeat. Forced join. Gap size estimated to
                be approximately 300bp by restriction digest data."
                complement(85021..87466)
                /locus_tag="RP11-503C24.5-001"
                /note="match: proteins: O15320 Q95046 Q96RT6 Q96SG9"
                /pseudocodon
                complement(85021..87466)
                /codon_start=1
                complement(join(98299..98705,100193..100230))
                /locus_tag="RP11-503C24.6-001"
                complement(join(98299..98705,100193..100230))
                /locus_tag="RP11-503C24.6-001"
                /note="match: ESTs: BC428602"
                complement(105020..117838)
                /gene="RP11-503C24.7"
                /locus_tag="RP11-503C24.7-001"
                complement(join(105020..107214,108284..108562,
                109268..109400,117504..117838))
                /gene="RP11-503C24.7"
                /locus_tag="RP11-503C24.7-001"
                /note="match: ESTs: AW150833 AW763170 BB638300 BB651889
                B1052777
                match: CDNA: AF318336"
                complement(105020)
                /gene="RP11-503C24.7"
                /locus_tag="RP11-503C24.7-001"
                complement(105046..105051)
                /gene="RP11-503C24.7"
                /locus_tag="RP11-503C24.7-001"
                complement(join(105548..107214,108284..108562,
                109268..109400,117504..117749))
                /gene="RP11-503C24.7"
                /locus_tag="RP11-503C24.7-001"
                /standard_name="OTTHUMP0000017658"
                /note="match: proteins: Q8WY22"
                /codon_start=1
                /protein_id="CA114850.1"
                /db_xref="GI:55962022"
                /db_xref="UniProt/TREMBL:Q5SN24"
                /translation="MWTGGPPGSGAGMDRRRLGARLRRAFAGLQSLQGLRATQOEKVR
                GALALOPPARPACGPHGLHGRBOELALALQSLRQLQDGLTTHLDLDLQ
                ISKLDDVGTASGALDSDSRBSGSGYEMSGSGSLSTSCASVCGDH1SP1AGSLP
                VAQAKKARSKMDKRRSYDRTTVPAKRRQATREBARPPGSGYEDAAQPMGCTPWPYVS
                TDDLRALPADTGLQASADAEILGLCGGVQIPLHVPKTRQDILVSGGABVYIP
                SP1HAVLQSPLEVLTKETPPRGSGSPFSPBSRGPAGLMTIGTQVLEAGPARAAYI
                DDLHLHMGRETPAKSGRGCGPLRLHAASPSROCGSWTDGGRLLVPAFGREDEGP
                AOSRGRGSGTPOOGYMPLEGGQSLPEBGSKPSNSCVLRRTMVAQSPSSAAQOTP
                SNADYGRGNIITSPSLDKSPSPASGHFAHPSGPAISLKKGPKPSKXAKTKRSPMDVL
                RRAQPLILDLRPSGAAHAPQPSLRENDPAHMTGCGGLQRRPALAWEACRCSBETL
                YPMFVLVPLAVALPQESHRTSAQLPFEASLILTVAKRRHMQSTVE1SAKARLASG
                PSSNIGPPPRVARRAAGPLARGPSLVRODALTYSDEPSKISAECDPFPVITPETS
                BESSDHTNTRRGRESSSDEGGQSGSCDCLALGCVAAAGHAEILAMTQEARVSSGGL
                LSPVPLKCRKIKASKKALKKKIRRFQPTALKVMTMV"
                112915
              misc_feature
```

misc_feature 115621 /note="Clone_right_end: RP11-3C9"
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."
misc_feature 159864
misc_feature 161496 /note="Clone_left_end: RP1-392"
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 420bp by restriction digest data."
ORIGIN
Query Match 83.2%; Score 20.8; DB 8; Length 161863;
Best Local Similarity 91.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATGTGACA 24
Db 135962 AAAAAAAAAAGCATGATGTGACA 135939
RESULT 28
AC069536 170941 bp DNA linear PRI 20-APR-2002
LOCUS Homo sapiens chromosome 10 clone RP11-112J5, complete sequence.
DEFINITION AC069536
ACCESSION AC069536.4 GI:20219055
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 170941)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170941)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 170941)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 170941)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Apr 20, 2002 this sequence version replaced gi:9887578.
FEATURES
source location/Qualifiers
1..170941
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-112J5"
/clone_1db="RP1-112J5"
ORIGIN
Query Match 83.2%; Score 20.8; DB 8; Length 170941;
Best Local Similarity 91.7%; Pred. No. 2.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATGTGACA 24
Db 135634 AAAAAAAAAAGCATGATGTGACA 135657
RESULT 29

AC018740/c
LOCUS 172823 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-449M6, WORKING DRAFT SEQUENCE,
16 unordered pieces.
ACCESSION AC018740
VERSION AC018740.3 GI:8570198
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 172823)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172823)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 17, 2000 this sequence version replaced gi:7024035.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0449M06
----- Summary Statistics -----
Sequencing vector: p13; 71%
Sequencing vector: plasmid; 29%
Chemistry: Dye-terminator Big Dye; 29% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 16631 bases at least Q40
Consensus quality: 168579 bases at least Q30
Consensus quality: 169265 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 171323; sum-of-contigs
Quality coverage: 6.09 in Q20 bases; sum-of-contigs
Quality coverage: 6.45 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2233: contig of 2233 bp in length
* 2234 2233: gap of unknown length
* 2334 5196: contig of 2863 bp in length
* 5197 5296: gap of unknown length
* 5297 9808: contig of 4512 bp in length
* 9809 9908: gap of unknown length
* 9909 15250: contig of 5342 bp in length
* 15251 15350: gap of unknown length
* 15351 20199: contig of 4849 bp in length
* 20200 20299: gap of unknown length
* 20300 25461: contig of 5162 bp in length
* 25462 25561: gap of unknown length
* 25562 34198: contig of 8637 bp in length
* 34199 34298: gap of unknown length
* 34299 43326: contig of 9028 bp in length
* 43327 43426: gap of unknown length
* 43427 51993: contig of 8567 bp in length
* 51994 52093: gap of unknown length
* 52094 64070: contig of 11977 bp in length
* 64071 64170: gap of unknown length
* 64171 78325: contig of 14155 bp in length

```
* 78326 78425: gap of unknown length
* 78426 93259: contig of 14834 bp in length
* 93260 93359: gap of unknown length
* 93360 106917: contig of 13558 bp in length
* 106918 107017: gap of unknown length
* 107018 125374: contig of 18357 bp in length
* 125375 125474: gap of unknown length
* 125475 145853: contig of 20379 bp in length
* 145854 145953: gap of unknown length
* 145954 172823: contig of 26870 bp in length.
```

FEATURES

```
source
1..172823
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-449M6"
1..223
/note="assembly_name:Contig21"
clone_end:T7
vector_side:right"
2234..2333
/estimated_length=unknown
2334..5196
/note="assembly_name:Contig22"
5197..5296
/estimated_length=unknown
5297..9808
/note="assembly_name:Contig23"
9809..9908
/estimated_length=unknown
9909..15250
/note="assembly_name:Contig24"
15251..15350
/estimated_length=unknown
15351..20199
/note="assembly_name:Contig25"
20200..20299
/estimated_length=unknown
20300..25461
/note="assembly_name:Contig26"
25462..25561
/estimated_length=unknown
25562..34198
/note="assembly_name:Contig27"
34199..34298
/estimated_length=unknown
34299..43326
/note="assembly_name:Contig28"
43327..43426
/estimated_length=unknown
43427..51593
/note="assembly_name:Contig29"
clone_end:SP6
vector_side:right"
51594..52093
/estimated_length=unknown
52094..64070
/note="assembly_name:Contig30"
64071..64170
/estimated_length=unknown
64171..78325
/note="assembly_name:Contig31"
78326..78425
/estimated_length=unknown
78426..93259
/note="assembly_name:Contig32"
93260..93359
/estimated_length=unknown
93360..106917
/note="assembly_name:Contig33"
106918..107017
/estimated_length=unknown
```

```
misc_feature 107018..125374
/note="assembly_name:Contig34"
gap 125375..125474
/estimated_length=unknown
misc_feature 125475..145853
/note="assembly_name:Contig35"
gap 145854..145953
/estimated_length=unknown
misc_feature 145954..172823
/note="assembly_name:Contig36"
```

ORIGIN

```
Query Match 83.2%; Score 20.8; DB 14; Length 172823;
Best Local Similarity 91.7%; Pred. No. 2,6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
Db 18742 AAAAAAAAAAGCATATTGTGCCA 18719
```

```
RESULT 30
AC090574/C 172969 bp DNA linear PRI 06-SEP-2002
LOCUS AC090574
DEFINITION Homo sapiens chromosome 8, clone RP11-449M6, complete sequence.
ACCESSION AC090574
VERSION AC090574.6 GI:22123296
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 172969)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-449M6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172969)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguski,M., Brown,A.,
Burt,D., Butler,A., Busch,C., Cannon,M., Cawthon,C.,
Cawthon,K., Cheng,K., Cho,K., Cline,D., Cohen,B.,
Cotton,M., D'Amico,A., DeLuca,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fero,S., Ferris,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagob,B., Heath,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Karas,L., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Margulis,N., Matthews,C., McCarthy,M., McKernan,K.,
McPherson,R., Meldrum,U., Menus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santor,R., Schauer,S., Schumacher,K., Seaman,S., Severy,P.,
Souarez,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zemke,L., Zimmer,A. and Zody,M.
```

TITLE Direct Submission
JOURNAL Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 172969)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Bloom,T., Boguski,M., Brown,A.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagob,B.,
Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C., Kemat,A.,
Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,O., Matthews,C.,
McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V.,


```

Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 172969)
Birten, B., Nushbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouknight, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:21431143.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11720
Center clone name: 449_M_6
-----
FEATURES
source
-----
location/Qualifiers
1. 172969
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-449M6"
/clone_lib="RC11 Human Male BAC"
726..746
/rpt_family="GC_rich"
3174..3325
/rpt_family="MIR"
4250..4444
/rpt_family="LIME"
complement(4481..4789)
/rpt_family="AluSg"
5002..5301
/rpt_family="AluSg"
5450..5470
/rpt_family="AluUo"
/rpt_family="AT_rich"
complement(6159..6407)
/rpt_family="MER112"
complement(6472..6555)
/rpt_family="MERSA"
6589..6670
/rpt_family="MERSB"
/rpt_family="MERSB"

repeat_region 6726..6841
/rpt_family="AluSg/x"
repeat_region 6843..7150
/rpt_family="AluSc"
repeat_region 7207..7509
/rpt_family="AluY"
complement(7990..8038)
/rpt_family="MER46B"
8206..8511
/rpt_family="AluSx"
complement(8583..8876)
/rpt_family="AluSx"
8900..9093
/rpt_family="LIME"
9094..9262
/rpt_family="MER46B"
9263..9488
/rpt_family="AluSg"
9489..9703
/rpt_family="MER46B"
11361..11612
/rpt_family="LIMC5"
complement(11614..11908)
/rpt_family="AluUo"
complement(11919..12216)
/rpt_family="AluSg"
12354..12436
/rpt_family="LIMC5"
12459..12766
/rpt_family="AluSc"
12774..13067
/rpt_family="AluSx"
13102..13122
/rpt_family="(TAA)n"
14397..14697
/rpt_family="AluSg"
14766..15068
/rpt_family="AluY"
complement(15083..15189)
/rpt_family="L2"
complement(15191..15389)
/rpt_family="AluUo"
15393..15561
/rpt_family="AluSg/x"
complement(16080..16165)
/rpt_family="MIR"
16227..16526
/rpt_family="AluUo"
17052..17200
/rpt_family="AluSg/x"
17215..17512
/rpt_family="AluSx"
17514..17750
/rpt_family="MIR"
complement(17935..18142)
/rpt_family="MER3"
18339..18661
/rpt_family="AluSx"
18682..18704
/rpt_family="(GAA)n"
18791..18906
/rpt_family="MERSA"
complement(18908..19127)
/rpt_family="AluUo"
complement(19136..19445)
/rpt_family="AluSx"
1956..20261
/rpt_family="AluUo"
20354..20577
/rpt_family="AluSx"
complement(20663..20789)
/rpt_family="AluUo"
20864..20887

```

```

repeat_region      /rpt family="AT_rich"
                    20868..21047
                    /rpt_family="FAM"
repeat_region      21048..21081

Query Match          83.2%; Score 20.8; DB 6; Length 172969;
Best Local Similarity 91.7%; Pred. No. 2,6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGCATGATTGTGACA 24
        |||||
Db       11928 AAAAAAAAAAGCATTAATTGTGCCA 11905

RESULT 31
LOCUS     AEO14835              180450 bp    DNA         linear   INV 07-OCT-2002
DEFINITION Plasmodium falciparum 3D7 chromosome 10 section 7 of 7 of the
complete sequence.
ACCESSION AEO14835 AEO14185
VERSION   AEO14835.1 GI:23495241
KEYWORDS
SOURCE    Plasmodium falciparum 3D7
ORANISM   Plasmodium falciparum 3D7
REFERENCE 1 (bases 1 to 180450)
AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
          Carlton,J.M., Paine,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
          James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
          Kyes,S., Chan,M.-S., Nene,V., Shalimov,S.J., Suh,B., Peterson,T.,
          Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
          Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
          Frumholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
          Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
          Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
          Barrell,B.
TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL   Nature 419 (6906), 498-511 (2002)
PUBMED   12368864
REFERENCE 2 (bases 1 to 180450)
AUTHORS   Direct Submission
          Gardner,M.J.
TITLE     Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
          Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source           1..180450
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="Genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="10"
                     complement(101..176)
     repeat_region    /rpt type=randem
                     /rpt_unit="(ta)n"
                     /rpt_length=101..165
     repeat_region    /rpt type=randem
                     /rpt_length=134..167
     repeat_region    /rpt type=randem
                     complement(184..209)
                     /note="AT_rich"
     repeat_region    /rpt type=randem
                     /rpt_length=257..283
     repeat_region    /rpt type=randem
                     complement(257..283)
     repeat_region    /rpt type=randem
                     /rpt_unit="(ta)n"
                     complement(646..677)
     repeat_region    /rpt type=randem
                     /rpt_unit="(a)n"
                     complement(687..721)
                     /note="AT_rich"
     repeat_region    /rpt type=randem
                     complement(729..823)

```

repeat_region	/rpt_type=tandem /rpt_unit="(taaaa)n" complement(880...1050)
repeat_region	/rpt_type=tandem /rpt_unit="(ta)n" 886...937
repeat_region	/rpt_type=tandem 993...1048
repeat_region	/rpt_type=tandem complement(1171... /note="Ar rich"
repeat_region	/rpt_type=tandem complement(1428... /rpt_type=tandem /rpt_unit="(ca)n" 1428...1464
repeat_region	/rpt_type=tandem complement(1519... /rpt_type=tandem /rpt_unit="(taaa)n" complement(1745...1778)
repeat_region	/rpt_type=tandem /rpt_unit="(ca)n" 1745...1774
repeat_region	/rpt_type=tandem complement(1896... /note="Ar rich"
repeat_region	/rpt_type=tandem complement(1999... /rpt_type=tandem /rpt_unit="(ta)n" complement(2164...2298)
repeat_region	/rpt_type=tandem /rpt_unit="(a)n" complement(2452...2492)
repeat_region	/rpt_type=tandem /rpt_unit="(taaa)n" 2468...2492
repeat_region	/rpt_type=tandem complement(2691... /note="Ar rich"
repeat_region	/rpt_type=tandem complement(2869... /rpt_type=tandem /rpt_unit="(ta)n" complement(3000...3095)
repeat_region	/rpt_type=tandem complement(3136... /rpt_type=tandem /rpt_unit="(ta)n" complement(3367... /note="Ar rich"
repeat_region	/rpt_type=tandem complement(3475... /note="Ar rich"
repeat_region	/rpt_type=tandem complement(3668... /note="Ar rich"
repeat_region	/rpt_type=tandem 3794...3826
repeat_region	/rpt_type=tandem complement(3794... /rpt_type=tandem /rpt_unit="(ta)n" complement(3842...3867)
repeat_region	/note="Ar rich"
repeat_region	/rpt_type=tandem 3884...3920
repeat_region	/rpt_type=tandem /rpt_unit="(a)n" complement(4023...4118)

```

/rpt_type=tandem      complement(4131. .4151)
repeat_region         /note="AT_rich"
/rpt_type=tandem      4383. .4409
repeat_region         /rpt_type=tandem complement(4383. .4409)
/rpt_type=tandem      /rpt_unit="(ca)n"
repeat_region         complement(4453. .4600)
/note="AT_rich"
/rpt_type=tandem      4615. .4653
repeat_region         /rpt_type=tandem complement(4615. .4653)
/rpt_type=tandem      /rpt_unit="(ca)n"
repeat_region         complement(4772. .4799)
/note="AT_rich"
/rpt_type=tandem      complement(4816. .4851)
/note="AT_rich"
repeat_region         /rpt_type=tandem complement(4931. .4981)
/rpt_type=tandem      /rpt_unit="(ca)n"
repeat_region         complement(4999. .5250)
/note="AT_rich"
/rpt_type=tandem      complement(5283. .5306)
/rpt_type=tandem      /rpt_unit="(ca)n"
repeat_region         complement(5348. .5381)
/note="AT_rich"
/rpt_type=tandem      5508. .5579
repeat_region         /rpt_type=tandem complement(5508. .5579)
/rpt_type=tandem      /rpt_unit="(ca)n"
repeat_region         complement(5583. .5611)
/rpt_type=tandem      complement(5583. .5611)
repeat_region         /rpt_type=tandem complement(5621. .5655)
/note="AT_rich"
/rpt_type=tandem      5656. .5689
repeat_region         /rpt_type=tandem complement(5730. .5756)
/rpt_type=tandem      /rpt_unit="(a)n"
repeat_region         /rpt_type=tandem complement(5834. .5861)
/rpt_type=tandem      /rpt_unit="(a)n"
repeat_region         complement(5929. .5956)
/note="AT_rich"
/rpt_type=tandem      complement(5957. .5985)
/rpt_type=tandem      /rpt_unit="(caaaa)n"
repeat_region         complement(6012. .6047)
/rpt_type=tandem      /rpt_unit="(a)n"
repeat_region         complement(6115. .6158)
/note="AT_rich"

```

Query Match 83.2%; Score 20.8; DB 2; Length 180450;
 Best Local Similarity 91.7%; Pred. No. 2.6e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 AAAAAAAAAACGATGTTGACAC 25
 Db 71134 AAAAAAAAAACGATGTTGATGAC 71157

RESULT 32

AC084303

LOCUS

DEFINITION

AC084303

VERSION

AC084303.2

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Homo.

1 (bases 1 to 181137)

Homo sapiens chromosome 10, clone RP11-344E7

Unpublished

2 (bases 1 to 181137)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,

Boukhalter, B., Brown, A., Burkett, G., Campolano, A., Castle, A.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,

Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,

Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,

Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K.,

Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,

Mackdonald, P., Marguis, N., McCarthy, M., McKernan, K.,

McPheters, R., Meldrum, J., Menes, L., Mihova, T., Menga, V.,

Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,

O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K.,

Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,

Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,

Souarez, C., Subramanian, A., Talama, J., Teste, S., Theodore, J.,

Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viet, R., Vo, A.,

Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,

Zimmer, A. and Zody, M.

Direct Submission

Submitted (12-OCT-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 181137)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,

Boukhalter, B., Brown, A., Burkett, G., Campolano, A., Castle, A.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,

Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,

Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,

Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K.,

Lamaze, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,

Mackdonald, P., Marguis, N., McCarthy, M., McKernan, K.,

McPheters, R., Meldrum, J., Menes, L., Mihova, T., Menga, V.,

Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,

O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K.,

Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,

Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,

Souarez, C., Subramanian, A., Talama, J., Teste, S., Theodore, J.,

Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viet, R., Vo, A.,

Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,

Zimmer, A. and Zody, M.

Direct Submission

Submitted (12-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 22, 2000 this sequence version replaced gi:10945764.

TITLE
 JOURNAL
 COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10560

Center clone name: 344.E.7

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 175856 bases at least Q40
 Consensus quality: 178964 bases at least Q30
 Consensus quality: 179890 bases at least Q20
 Insert size: 182000; agarose-fp
 Insert size: 180357; sum-of-contrigs
 Quality coverage: 5.5 in Q20 bases; agarose-fp
 Quality coverage: 5.6 in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contrigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      584: contrig of 584 bp in length
*      585      684: gap of 100 bp
*      685      2288: contrig of 1604 bp in length
*      2289      2388: gap of 100 bp
*      2389      4654: contrig of 2266 bp in length
*      4655      4754: gap of 100 bp
*      4755      87524: contrig of 82770 bp in length
*      87525      87624: gap of 100 bp
*      87625      97410: contrig of 9786 bp in length
*      97411      97510: gap of 100 bp
*      97511      107863: contrig of 10353 bp in length
*      107864      107963: gap of 100 bp
*      107964      124764: contrig of 16801 bp in length
*      124765      124864: gap of 100 bp
*      124865      176131: contrig of 51267 bp in length
*      176132      176232: gap of 100 bp
*      176232      181137: contrig of 4906 bp in length.

```

FEATURES

source

```

1..181137
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10"
/clone="RP11-344E7"
/clone_1lb="RP11-344E7"
1..584
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
585..684
/estimated_length=100
685..2288
/note="assembly_fragment"
2289..2388
/estimated_length=100
2389..4654
/note="assembly_fragment"
4655..4754
/estimated_length=100
4755..87524
/note="assembly_fragment"

```

ORIGIN

```

gap      87525..87624
/estimated_length=100
misc_feature      87625..97410
/note="assembly_fragment"
gap      97411..97510
/estimated_length=100
misc_feature      97511..107863
/note="assembly_fragment"
gap      107864..107963
/estimated_length=100
misc_feature      107964..124764
/note="assembly_fragment"
gap      124765..124864
/estimated_length=100
misc_feature      124865..176131
/note="assembly_fragment"
gap      176132..176231
/estimated_length=100
misc_feature      176232..181137
/note="assembly_fragment"
clone_end:77
vector_side:right"

```

```

Query Match      83.2%; Score 20.8; DB 14; Length 181137;
Best Local Similarity 91.7%; Pred. No. 2.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAGCATGTTGTGACA 24
Db      14721 AAAAAAAAAAGCATGTTGTGCA 14744

```

```

RESULT 33      AC084251      181400 bp      DNA      linear      PRI 28-SEP-2002
LOCUS      AC084251
DEFINITION      Homo sapiens chromosome 8, clone RP11-152C15, complete sequence.
ACCESSION      AC084251
VERSION      AC084251.13 GI:23343878
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens

```

```

REFERENCE      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
TITLES      Homnidae; Homo.
JOURNAL      1 (bases 1 to 181400)
REFERENCE      Birren, B., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
AUTHORS      Birren, B., Linton, L., Bastien, V., Bada, F., Boguslavsky, L.,
TITLES      Anderson, S., Barna, N., Bastien, V., Bada, F., Boguslavsky, L.,
JOURNAL      Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
AUTHORS      Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
TITLES      DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
JOURNAL      Fitzhugh, W., Gage, D., Galagan, J., Gardy, S., Ginde, S., Goyette, M.,
AUTHORS      Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
TITLES      Iliev, I., Johnson, R., Jones, C., Kann, L., Karalae, A., LaRoque, K.,
JOURNAL      Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
AUTHORS      Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McKernan, K.,
TITLES      McPheters, R., Melidrim, J., Menes, J., Minova, T., Mlenga, V.,
JOURNAL      Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
AUTHORS      O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K.,
TITLES      Pierre, N., Pisanil, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
JOURNAL      Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
AUTHORS      Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
TITLES      Strauss, N., Subramanian, A., Talame, J., Testfaye, S., Theodore, J.,
JOURNAL      Tittel, A., Travers, W., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,
AUTHORS      Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
TITLES      Zimmer, A. and Zody, M.
JOURNAL      Direct Submissions

```

Submitted (18-OCT-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

```

REFERENCE
AUTHORS
3 (bases 1 to 181400)
Birren, B., Nusbaum, C., Lander, E. A., Allen, N., Anderson, S.,
Batra, N., Baetien, V., Bloom, T., Boguslavsky, L., Bouknight, B.,
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hages, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Miengo, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Sciojanovic, N., Talamas, J.,
Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
Direct Submission
Submitted (08-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 181400)
Birren, B., Nusbaum, C., Lander, E. A., Allen, N., Anderson, S.,
Batra, N., Baetien, V., Bloom, T., Boguslavsky, L., Bouknight, B.,
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hages, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Miengo, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Sciojanovic, N., Talamas, J.,
Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
Direct Submission
Submitted (28-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 28, 2002 this sequence version replaced gi:22758796.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11326
Center clone name: 152_C_15

-----
FEATURES
SOURCE
1. 181400
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-152C15"
/clone_11b="RPC1-11 Human Male BAC"
12. 45
/rpt_family="(A)n"
repeat_region
46. 768
/rpt_family="(TTTGG)n"
repeat_region
769. 1080
/rpt_family="(AluSp)"
repeat_region
1081. 1138
/rpt_family="(AluSp)"
repeat_region
/rpt_family="(AT_rich"
complement(1166. .1319)
/rpt_family="(AluIo)"
complement(1320. .1619)
/rpt_family="(AluY)"
complement(1620. .1750)
/rpt_family="(AluIo)"
complement(1767. .2068)
/rpt_family="(AluSg)"
2084. 2261
/rpt_family="(AluY)"
complement(2495. .2802)
/rpt_family="(AluSg)"
complement(3647. .3911)
/rpt_family="(AluY)"
complement(3913. .4021)
/rpt_family="(AluIo/FLAM"
4071. 4365
/rpt_family="(AluSx)"
complement(4780. .5083)
/rpt_family="(AluY)"
complement(5334. .5648)
/rpt_family="(AluSx)"
complement(5649. .5751)
/rpt_family="(MIR"
5806. 5952
/rpt_family="(AluY)"
6005. 6062
/rpt_family="(AluY/FLAM"
complement(6091. .6304)
/rpt_family="(L2"
6305. 6568
/rpt_family="(AluS)"
complement(6569. .6772)
/rpt_family="(L2"
6806. 6992
/rpt_family="(MIR"
complement(7113. .7276)
/rpt_family="(MIR"
10227. 10251
/rpt_family="(A)n"
10322. 10391
/rpt_family="(A-rich"
10649. 10836
/rpt_family="(MERSA"
10879. 11033
/rpt_family="(MERSA"
complement(11603. .11908)
/rpt_family="(AluSg)"
11940. 11971
/rpt_family="(AT_rich"
12417. 12717
/rpt_family="(AluSg)"
complement(12973. .13263)
/rpt_family="(AluY)"
complement(13281. .13594)
/rpt_family="(AluSx)"
14122. 14415
/rpt_family="(AluSx)"
14428. 14746
/rpt_family="(AluSg)"
complement(14922. .15219)
/rpt_family="(AluY)"
15296. 15318
/rpt_family="(TTTA)n"
15348. 15374
/rpt_family="(TTTTGG)n"
complement(15390. .15698)
/rpt_family="(AluSp)"
complement(15699. .15974)
/rpt_family="(AluSg)"

```

```

repeat_region complement(15975..16110)
/rpc_family="AlusG/x"
repeat_region 16127..16247
/rpc_family="L1MB6"
repeat_region complement(16248..16542)
/rpc_family="AlusX"
repeat_region 16543..16897
/rpc_family="L1MB6"
repeat_region 16898..17207
/rpc_family="AlusP"
repeat_region 17208..17250
/rpc_family="L1MB6"
repeat_region 17251..17313
/rpc_family="AlusB"
repeat_region 17546..17574
/rpc_family="AT_rich"

Query Match 83.2%; Score 20.8; DB 8; Length 181400;
Best Local Similarity 91.7%; Pred. No. 2.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 90625 AAAAAAAAAAGCATGATTGTGACA 90648

RESULT 34
AC133942 186014 bp DNA linear ROD 28-JUL-2004
LOCUS Mus musculus chromosome 7 clone RP23-472J11, complete sequence.
DEFINITION AC133942
AC133942
VERSION AC133942.3 GI:50727091
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Wilson, R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 186014)
McPherson, J.D. and Waterston, R.H.
TITLE Direct Submision
JOURNAL Submitted (20-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 186014)
Wilson, R.K.
TITLE Direct Submision
JOURNAL Submitted (08-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 186014)
Wilson, R.K.
TITLE Direct Submision
JOURNAL Submitted (28-JUL-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 28, 2004 this sequence version replaced gi:47084589.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
Project Information
----- Project Information -----
Center project name: M BA0472J11
-----

FEATURES
source location/Qualifiers
1..186014
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="7"

```

```

ORIGIN
/clone="RP23-472J11"

Query Match 83.2%; Score 20.8; DB 9; Length 186014;
Best Local Similarity 91.7%; Pred. No. 2.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 13515 AAAAAAAAAAGCATGATTGTGACA 13538

RESULT 35
BX530070 191146 bp DNA linear VRT 29-JAN-2004
LOCUS Zebrafish DNA sequence from clone DKEY-245C12 in linkage group 7,
DEFINITION BX530070
BX530070
VERSION BX530070.9 GI:40994813
KEYWORDS HTG.
SOURCE Danto rerio (zebrafish)
ORGANISM Danto rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 191146)

REFERENCE
AUTHORS Wray, P.
TITLE Direct Submision
JOURNAL Submitted (15-JAN-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
On Jan 17, 2004 this sequence version replaced gi:38304125.

COMMENT
----- Genome Center -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
-----

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats, where
this is found the longest good quality representation will be
submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhifeng Bao and Sean Eddy, submitted), and those
beginning 'dir' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-245C12
is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.

FEATURES
source location/Qualifiers
1..191146
/organism="Danto rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"

```

ORIGIN /clone="DKEY-245C12"
/clone_lib="DanticKey"

Query Match 83.2% Score 20.8; DB 5; Length 191146;
Best Local Similarity 91.7% Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 113978 AAAAAAAAAAGCATGAGTGACA 114001

RESULT 36
AC098722 197946 bp DNA linear ROD 31-OCT-2003
LOCUS Mus musculus BAC clone RP23-3B20 from 14, complete sequence.
DEFINITION AC098722
ACCESSION AC098722.3 GI:19909473
VERSION
KEYWORDS
SOURCE HTG.
MUS musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Vanbrunt, A. and Haglund, K.
TITLE The sequence of Mus musculus BAC clone RP23-3B20
JOURNAL Unpublished (2001)
REFERENCE
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2001) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 197946)
McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-ARR-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 197946)
McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 197946)
Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 3, 2002 this sequence version replaced gi:16924167.

REFERENCE
AUTHORS
JOURNAL
COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
Summary Statistics
Center project name: M_BA0003B20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence

from more than one subclone, and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES

source	Location/Qualifiers
1..197946	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/DB_xref="taxon:10090"
	/chromosome="14"
	/map="14"
	/clone="RP23-3B20"
	/clone_lib="RPCI-23"
repeat_region	266..515 /rpt_family="L1"
repeat_region	2792..3015 /rpt_family="B4"
repeat_region	3260..4096 /rpt_family="L1"
repeat_region	4255..4441 /rpt_family="ERVY"
repeat_region	4950..5035 /rpt_family="ERVY"
repeat_region	5173..5309 /rpt_family="Alu"
repeat_region	5375..5557 /rpt_family="L1"
repeat_region	5628..5924 /rpt_family="L1"
repeat_region	6431..6498 /rpt_family="ERV1"
repeat_region	9060..9274 /rpt_family="L1"
repeat_region	9470..9710 /rpt_family="MALR"
repeat_region	9711..9761 /rpt_family="ERV1"
repeat_region	11576..12015 /rpt_family="MALR"
repeat_region	12386..12541 /rpt_family="L1"
repeat_region	12542..12729 /rpt_family="B2"
tRNA	complement(12652..12723) /product="tRNA-Asn"
	/note="Likely pseudogene (HMM Sc=33.75 / Sec struct Sc=-10.38)"
repeat_region	12730..12818 /rpt_family="L1"
repeat_region	15038..15411 /rpt_family="L1"
repeat_region	15400..15970 /rpt_family="L1"
repeat_region	15980..16240 /rpt_family="MALR"
repeat_region	16962..17206 /rpt_family="B4"
repeat_region	17534..17570

```

repeat_region      /rpt_family="L2"
                    17697..18097
repeat_region      /rpt_family="MALR"
                    20083..20542
repeat_region      /rpt_family="L1"
                    22621..22659
repeat_region      /rpt_family="L1"
                    22675..22774
repeat_region      /rpt_family="L1"
                    26597..26754
repeat_region      /rpt_family="L1"
                    26851..26909
repeat_region      /rpt_family="ERV1"
                    28126..28898
repeat_region      /rpt_family="L1"
                    28938..30079
repeat_region      /rpt_family="L1"
                    30079..30302
repeat_region      /rpt_family="L1"
                    30302..30598
repeat_region      /rpt_family="L1"
                    30579..30688
repeat_region      /rpt_family="L1"
                    30689..30915
repeat_region      /rpt_family="MER1_type"
                    30916..31004
repeat_region      /rpt_family="L1"
                    31230..31329
repeat_region      /rpt_family="B2"
                    32359..33243
repeat_region      /rpt_family="L1"
                    34189..35519
repeat_region      /rpt_family="L1"
                    35522..35640
repeat_region      /rpt_family="L1"
                    38119..38301
repeat_region      /rpt_family="L1"
                    38561..38925
repeat_region      /rpt_family="L1"
                    39008..39335
repeat_region      /rpt_family="L1"
                    40326..41957
repeat_region      /rpt_family="L1"
                    41873..42262
repeat_region      /rpt_family="L1"
                    42320..42493
repeat_region      /rpt_family="B2"
                    42569..42767
repeat_region      /rpt_family="L1"
                    44906..45383
repeat_region      /rpt_family="L1"
                    45383..45568
repeat_region      /rpt_family="L1"
                    47234..47434
repeat_region      /rpt_family="L1"
                    47674..47811
repeat_region      /rpt_family="L1"
                    48411..48444
repeat_region      /rpt_family="U2"
                    49076..49212
repeat_region      /rpt_family="Alu"
                    49894..50155
repeat_region      /rpt_family="RMRB12"
                    50206..50586
repeat_region      /rpt_family="RMRB12"
                    51144..51200
repeat_region      /rpt_family="ERV1"
                    53580..53431
repeat_region      /rpt_family="L1"

```

```

Query Match      83.2%; Score 20.8; DB 9; Length 197946;
Best Local Similarity 91.7%; Pred.No.2.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 AAAAAAAAAAGCATGTGTGACA 24
Db      71136 AAAAAACAAAGCATTTGTGACA 71159

RESULT 37
AC021185
LOCUS
DEFINITION
AC021185      201611 bp      DNA      11near      HTG 20-OCT-2001
Homo sapiens chromosome 6 clone RP11-73815, WORKING DRAFT SEQUENCE,
5 unordered pieces.
ACCESSION
AC021185
VERSION
AC021185.4 GI:16259198
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Homo sapiens
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 201611)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 201611)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Oct 18, 2001 this sequence version replaced gi:10944513.

COMMENT
----- Genome Center -----
Center: Washington University genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@waterston.wustl.edu
----- Project Information -----
Center project name: H_NH0738105
----- Summary Statistics -----
Sequencing vector: Plasmid; 35%
Chemistry: Dye-primer ET; 35% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 201095 bases at least Q40
Consensus quality: 201923 bases at least Q30
Consensus quality: 202374 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      16969: contig of 16969 bp in length
*      16970      17069: gap of unknown length
*      17070      33654: contig of 16585 bp in length
*      33655      33754: gap of unknown length
*      33755      53582: contig of 19828 bp in length
*      53583      53682: gap of unknown length
*      53683      126590: contig of 72908 bp in length
*      126591      126690: gap of unknown length
*      126691      201611: contig of 74921 bp in length.
*
Location/Qualifiers
1..201611
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-73815"

```



```
misc_feature 1. .16969
gap /note="assembly_name:Contig34
16970. 17069
/estimated_length=unknown
17070. .33654
misc_feature /note="assembly_name:Contig35
17070. .33654
gap /note="assembly_name:Contig36
17070. .33754
/estimated_length=unknown
33755. .53582
misc_feature /note="assembly_name:Contig36
clone_end:T7
vector_side:right"
gap 53583. .53682
/estimated_length=unknown
53683. .126390
misc_feature /note="assembly_name:Contig37
126390. .126390
gap /estimated_length=unknown
126691. .20611
misc_feature /note="assembly_name:Contig38
clone_end:SP6
vector_side:left"
```

```
FEATURES
      source
          Location/Qualifiers
            1..202495
              /organism="Papio anubis"
              /mol_type="genomic DNA"
              /db_xref="taxon:9555"
              /clone="RP41-28A19"
              /clone_11b="RP41"
              /note="BAC resource: http://bacpac.chori.org/"
      misc_feature
          1..54652
            /note="clone overlaps with GenBank Accession Number
            AC150737 clone RP41-467K18 (center project name hpe)"
      misc_feature
          1..3497
            /note="assembly_fragment
            clone_end:SP6
            vector_side:left"
      gap
          3498..3597
            /estimated_length=unknown
            /note="32031
            /note="assembly_fragment"
      misc_feature
          3598..32032
            /estimated_length=unknown
            /note="32131
            /estimated_length=unknown
      misc_feature
          32132..64401
            /note="assembly_fragment"
      gap
          64402..64501
            /estimated_length=unknown
      misc_feature
          64502..69214
            /note="assembly_fragment"
      gap
          69215..69314
            /estimated_length=unknown
      misc_feature
          69315..80359
            /note="assembly_fragment"
      gap
          80360..80459
            /estimated_length=unknown
      misc_feature
          80460..89883
            /note="assembly_fragment"
      gap
          89884..89983
            /estimated_length=unknown
      misc_feature
          89984..93768
            /note="assembly_fragment"
      gap
          93769..93868
            /estimated_length=unknown
      misc_feature
          93869..95660
            /note="assembly_fragment"
      gap
          95661..95760
            /estimated_length=unknown
      misc_feature
          95761..106536
            /note="assembly_fragment"
      gap
          106537..106636
            /estimated_length=unknown
      misc_feature
          106637..112578
            /note="assembly_fragment"
      gap
          112579..112678
            /estimated_length=unknown
      misc_feature
          112679..114908
            /note="assembly_fragment"
      gap
          114909..115008
            /estimated_length=unknown
      misc_feature
          115009..144349
            /note="assembly_fragment"
      gap
          144350..144449
            /estimated_length=unknown
      misc_feature
          144450..150821
            /note="assembly_fragment"
      gap
          150822..150921
            /estimated_length=unknown
      misc_feature
          150922..152838
            /note="assembly_fragment"
      gap
          152839..152938
            /estimated_length=unknown
      misc_feature
          152939..172089
            /note="assembly_fragment"
      gap
          172090..172189
            /estimated_length=unknown
```

```
      misc_feature
          172190..191169
            /note="assembly_fragment"
      gap
          191170..191269
            /estimated_length=unknown
      misc_feature
          191270..202495
            /note="assembly_fragment
            missing 17 clone end on 3' end of insert"
      misc_feature
          198135..202495
            /note="clone overlaps with GenBank Accession Number
            AC150736 clone RP41-405J6 (center project name hpg)"
      ORIGIN
          Query Match      83.2%; Score 20.8; DB 14; Length 202495;
          Best Local Similarity 91.7%; Pred. No. 2.5e+02;
          Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

      QY      1 AAAAAAAAAAGCATGATGTGACACA 24
      Db      68444 AAAAAAGAAAGCATGCTGTGACA 68421

      RESULT 39
      AC108435/c
      LOCUS      AC108435
      DEFINITION Mus musculus chromosome 7, clone RP23-229020, complete sequence.
      AC108435
      VERSION      AC108435.5 GI:32813518
      KEYWORDS      HTG.
      SOURCE      Mus musculus (house mouse)
      ORGANISM      Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridea; Muridae; Murinae; Mus.
      REFERENCE
          1 (bases 1 to 203281)
            AUTHORS      Birren, B., Nusbaum, C. and Lander, E.
            TITLE      Mus musculus chromosome 7, clone RP23-229020
            JOURNAL      Unpublished
            REFERENCE
          2 (bases 1 to 203281)
            AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
            Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
            Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B.,
            Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
            Cooke, P., Dearellano, K., Dewar, K., Diaz, D., S., Dodge, S., Fero, S.,
            Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
            Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
            Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
            Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
            Landers, T., Lehocsky, J., Levine, R., Liu, G., Maclean, C.,
            Macdonald, P., Major, V., Marquis, N., Matthews, C., McCarthy, M.,
            McEwan, P., McKernan, K., Meltrin, J., Menues, L., Mihova, T.,
            Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
            Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
            Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
            Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
            Rossetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,
            Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
            Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J.,
            Topham, K., Travers, M., Travis, N., Trifillo, J., Vassiliev, H.,
            Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W. J., Young, G.,
            Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 203281)
            AUTHORS      Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N.,
            Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, N.,
            Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
            Collymore, A., Cook, A., Cooke, P., Cornu, B., Dearellano, K.,
            Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,
            Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
            Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
            Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
            Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
```

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (10-JUN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 203281)
Birtten, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N.,
Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J. S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 203281)
Birtten, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N.,
Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J. S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 16, 2003 this sequence version replaced gi:30985001.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://fpc.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20764

Center clone name: 229_O_20

FEATURES	Location/Qualifiers
Source	1..203281
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="7"
	/map="7"
	/clone="RP23-229020"
	/clone_11b="RPCI-23 Female Mouse BAC"
misc_feature	1..5545
	/note="wgs end extension"
	clone_end:SpE"
repeat_region	2941..3026
	/rpt_family="(CACCT)n"
repeat_region	complement(3073..4156)
	/rpt_family="(Lx5"
repeat_region	4157..4186
	/rpt_family="(T)n"
repeat_region	complement(4218..6684)
	/rpt_family="(L1_MM"
misc_feature	5546..5551
	/note="clone boundary"
	clone_end:SpE
	Site:ECORI"
repeat_region	6826..6918
	/rpt_family="(B1F"
repeat_region	6921..6951
	/rpt_family="(CAA)n"
repeat_region	7198..7255
	/rpt_family="(TG)n"
repeat_region	complement(8578..9046)
	/rpt_family="(L1_MM"
repeat_region	9562..9607
	/rpt_family="(TTTG)n"
repeat_region	10222..10293
	/rpt_family="(PB1D10"
repeat_region	complement(10422..10551)
	/rpt_family="(B3"
repeat_region	10564..11090
	/rpt_family="(L1_MM"
repeat_region	11121..11200
	/rpt_family="(CGTG)n"
repeat_region	complement(11201..11358)
	/rpt_family="(B3A"
repeat_region	12488..12651
	/rpt_family="(RGTNE1"
repeat_region	complement(12944..13160)
	/rpt_family="(MTC"
repeat_region	complement(13208..13317)
	/rpt_family="(L1_MM"
repeat_region	complement(13318..13464)
	/rpt_family="(B2_MM2"
repeat_region	complement(13468..13683)
	/rpt_family="(MTC"
repeat_region	complement(13705..14021)
	/rpt_family="(ORR1D"
repeat_region	14120..14392
	/rpt_family="(RLTR14"
repeat_region	14393..14576
	/rpt_family="(B2_MM2"
repeat_region	14577..14671
	/rpt_family="(RLTR14"
repeat_region	14685..14833
	/rpt_family="(B1_MM"
repeat_region	14890..15020
	/rpt_family="(L1M2"
repeat_region	complement(15060..15113)
	/rpt_family="(MT2A"
repeat_region	15114..15140
	/rpt_family="(TCTA)n"
repeat_region	complement(15141..15567)

repeat_region /ftp_family="MT2A"
15576.15684
/db_family="LIM2"
repeat_region complement(18196..18487)

Query Match 83.2%; Score 20.8; DB 9; Length 203281;
Best Local Similarity 91.7%; Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
Db 6230 AAAAAAAAAATCATGATTGTGACA 6207

RESULT 40
BX510655 205761 bp DNA linear VRT 01-APR-2005
LOCUS Zebrafish DNA sequence from clone DKEY-94J16 in linkage group 5,
DEFINITION complete sequence.
ACCESSION BX510655 GI:62148892
VERSION BX510655.18 GI:62148892
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 205761)
Tracey, A.
Direct Submission
Submitted (01-APR-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf181-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 1, 2005 this sequence version replaced gi:61656758.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf181-help@sanger.ac.uk

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; SW; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish cDNA subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-94J16 is from a Zebrafish BAC library

FEATURES

source
1..205761
location/Qualifiers
/organism="Danio rerio"

ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 205761;
Best Local Similarity 91.7%; Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
Db 108802 AAAAAAAAAAGCATGATTGTTTACA 108825

RESULT 41
AC161795/c 208833 bp DNA linear HTG 05-JUL-2005
LOCUS Mus musculus chromosome 18 clone RP23-121G22 map 18.*** SEQUENCING
DEFINITION IN PROGRESS ***, 7 unordered pieces.
ACCESSION AC161795
VERSION AC161795.4 GI:68563551
KEYWORDS HTG; HTGS PHASE1; HTGS_FULTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 208833)
Birren, B., Nusbaum, C. and Lander, R.
Mus musculus chromosome 18, clone RP23-121G22
Unpublished
2 (bases 1 to 208833)

REFERENCE

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Anderson, M., Anderson, S., Arachchi, H.M., Barua, N., Baetien, V.,
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Bloom, T., Boguslavskiy, L., Boukhalter, B., Canatka, J., Chang, J.,
Choepe, Y., Collamore, A., Cook, A., Cooke, P., Corum, B.,
Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Menues, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norby, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhthang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
Reiter, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (20-MAY-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208833)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Anderson, S., Arachchi, H.M., Barua, N., Baetien, V.,
Bloom, T., Boguslavskiy, L., Boukhalter, B., Canatka, J., Chang, J.,
Choepe, Y., Collamore, A., Cook, A., Cooke, P., Corum, B.,
Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T.,
Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
McCarthy, M., Meldrim, J., Menues, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norby, C.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunhthang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,

Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N.,
Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J.,
Tophan, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R.,
Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
Zimmer, A., and Zody, M.
Submitted (05-JUL-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
On Jul 5, 2005 this sequence version replaced gi:66954872.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

TITLE
JOURNAL

Center: Broad Institute of MIT and Harvard
Center code: W1BR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@broad.mit.edu
Project Information
Center project name: L32528
Center clone name: 121_G_22

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 24346: contig of 24346 bp in length
* 24347 24446: gap of unknown length
* 24447 54069: contig of 29623 bp in length
* 54070 54169: gap of unknown length
* 54170 71006: contig of 16837 bp in length
* 71007 71106: gap of unknown length
* 71107 89106: contig of 18000 bp in length
* 89107 89206: gap of unknown length
* 89207 91845: contig of 2639 bp in length
* 91846 91945: gap of unknown length
* 91946 188249: contig of 96304 bp in length
* 188250 208833: gap of unknown length
* 208833: contig of 20484 bp in length.
Location/Qualifiers

FEATURES
source

1. 208833
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/map="18"
/clone="RP23-121G22"
/clone_lib="RP23 Female Mouse BAC"
24347..24446
/estimated_length=unknown
54070..54169
/estimated_length=unknown
71007..71106
/estimated_length=unknown
89107..89206
/estimated_length=unknown
91846..91945
/estimated_length=unknown
188250..188349
/estimated_length=unknown

ORIGIN

Query Match 83.2% Score 20.8; DB 14; Length 208833;
Best Local Similarity 91.7% Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATGTGACA 24
DB 87188 AAAAAAAAAAGCATGATGTGACA 87165

RESULT 42
AC163593/c
LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC163593 213024 bp DNA linear HTG 01-JUL-2005
Bos taurus clone CH240-12515, *** SEQUENCING IN PROGRESS ***, 22
unordered pieces.
AC163593.2 GI:68299943
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 213024)
Mazy, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D.,
Ayala-Bechl, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dedert, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falle, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, R., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabist, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gubaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hu, Y. S., Hume, J., Idelbrd, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowit, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenschwet, L., Loulesged, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindratne, M., Mahoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milojevic, S., Miner, G., Ming, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackelme, O., Okunolu, G., Olamugboon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfamkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Pitman, E., Pu, L., L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, R., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shetman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, B., Umanji, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, K., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 213024)
Worley, K. C.
Direct Submission
Submitted (13-JUN-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (baees 1 to 213024)
 AUTHORS Cow Genome Sequencing Consortium.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jun 29, 2005 this sequence version replaced gi:67514866.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: FHTB

Center clone name: CH240-12515

Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 203030 bases at least Q40

Consensus quality: 205410 bases at least Q30

Consensus quality: 207269 bases at least Q20

Estimated insert size: 206505; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docx/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 22 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

```

* 1 24745: contig of 24745 bp in length
* 24746 25290: gap of 545 bp
* 25291 57782: contig of 32492 bp in length
* 57783 57859: gap of 77 bp in length
* 57860 65020: contig of 7161 bp in length
* 65021 65206: gap of 186 bp
* 104117 104117: contig of 38911 bp in length
* 104118 104167: gap of 50 bp
* 132634 132634: contig of 28467 bp in length
* 132635 132684: gap of 50 bp
* 132685 133052: contig of 10368 bp in length
* 143053 143102: gap of 50 bp
* 143103 150638: contig of 7536 bp in length
* 150639 150688: gap of 50 bp
* 150689 156638: contig of 5950 bp in length
* 156639 156688: gap of 50 bp
* 156689 159769: contig of 13081 bp in length
* 159770 169819: gap of 50 bp
* 169820 174121: contig of 4302 bp in length
* 174122 174171: gap of 50 bp
* 184813 184813: contig of 10642 bp in length
* 184814 184863: gap of 50 bp
* 184864 191356: contig of 6493 bp in length
* 191357 191406: gap of 50 bp
* 191407 195160: contig of 3754 bp in length
* 195161 195509: gap of 349 bp
* 195510 198763: contig of 3254 bp in length
* 198764 198863: gap of unknown length

```

FEATURES

```

source
1..213024
  /organism="Bos taurus"
  /mol_type="genomic DNA"
  /db_xref="taxon:9913"
  /clone="CH240-12515"
  24746..25290
    /estimated_length=545
  57783..57859
    /estimated_length=77
  65021..65206
    /estimated_length=186
  104118..104167
    /estimated_length=50
  132635..132684
    /estimated_length=50
  143053..143102
    /estimated_length=50
  150639..150688
    /estimated_length=50
  156639..156688
    /estimated_length=50
  169770..169819
    /estimated_length=50
  174122..174171
    /estimated_length=50
  184814..184863
    /estimated_length=50
  191357..191406
    /estimated_length=50
  195161..195509
    /estimated_length=349
  198764..198863
    /estimated_length=unknown
  200174..200273
    /estimated_length=unknown
  201861..201960
    /estimated_length=unknown
  203231..203330
    /estimated_length=unknown
  204355..204454
    /estimated_length=unknown
  205503..205602
    /estimated_length=unknown
  207055..207154
    /estimated_length=unknown
  208252..208351
    /estimated_length=unknown
  213024..213024
    /estimated_length=unknown

```

```

Query Match 83.2%; Score 20.8; DB 14; Length 213024;
Best Local Similarity 91.7%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 AAAAAAAAAAGCATGATGTGACCA 24
Db 58879 AAAAAAAAAAGCATGACAGTGACA 58856

```

```

RESULT 43
AC161925 213666 bp DNA linear HTG 22-MAY-2005
LOCUS Mus musculus chromosome 18 clone Rp23-59L5 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 35 unordered pieces.

```

AC161925
 AC161925.1 GI:66392667
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 213666)
 Mus musculus chromosome 18, clone RP23-59L5
 Unpublished
 2 (bases 1 to 213666)
 Biren, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N.,
 Anderson, M., Andersson, S., Arachchi, H.M., Barina, N., Bastien, V.,
 Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J.,
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corm, B.,
 D'Arrellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
 Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D.,
 Galagan, J., Gardina, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Hagopian, D., Hagos, B., Hall, U., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T.,
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R.,
 MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C.,
 McCarthy, M., Meldrum, J., Meneus, L., Minova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkiang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
 Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomson, N.,
 Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Vassiliev, H., Venkataram, V.S., Viel, R.,
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (22-MAY-2005) Broad Institute of MIT and Harvard, 320
 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Broad Institute of MIT and Harvard
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@broad.mit.edu
 ----- Project Information
 Center project name: L32553
 Center clone name: 59.L.5
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 200447 bases at least Q40
 Consensus quality: 205025 bases at least Q30
 Consensus quality: 208388 bases at least Q20
 Insert size: 230000; agarose-fp
 Insert size: 210266; sum-of-contigs
 Quality coverage: 4.1 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 35 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1333: contig of 1333 bp in length
 * 1334 1433: gap of unknown length
 * 1434 2451: contig of 1018 bp in length
 * 2452 2552: gap of unknown length
 * 2552 3569: contig of 1018 bp in length

FEATURES

source

3570 3669: gap of unknown length
 3670 5895: contig of 2226 bp in length
 5896 5995: gap of unknown length
 5996 7002: contig of 1007 bp in length
 7003 7102: gap of unknown length
 7103 10154: contig of 3052 bp in length
 10155 10254: gap of unknown length
 10255 13202: contig of 2948 bp in length
 13203 13302: gap of unknown length
 13303 15343: contig of 2040 bp in length
 15344 15442: gap of unknown length
 15443 18266: contig of 2824 bp in length
 18267 18367: gap of unknown length
 18368 20817: contig of 2451 bp in length
 20818 20917: gap of unknown length
 20918 24056: contig of 3139 bp in length
 24057 24155: gap of unknown length
 24157 27612: contig of 3456 bp in length
 27613 27712: gap of unknown length
 27713 30777: contig of 3065 bp in length
 30778 30877: gap of unknown length
 30878 33607: contig of 2730 bp in length
 33608 33707: gap of unknown length
 33708 36632: contig of 2925 bp in length
 36633 36732: gap of unknown length
 36733 40506: contig of 3774 bp in length
 40507 40607: gap of unknown length
 40608 43166: contig of 2560 bp in length
 43167 43266: gap of unknown length
 43267 48355: contig of 5089 bp in length
 48356 48455: gap of unknown length
 48456 51702: contig of 3247 bp in length
 51703 51802: gap of unknown length
 51803 57203: contig of 5401 bp in length
 57204 57303: gap of unknown length
 57304 59979: contig of 2676 bp in length
 59980 60079: gap of unknown length
 60080 63103: contig of 3024 bp in length
 63104 63203: gap of unknown length
 63204 67272: contig of 4069 bp in length
 67273 67372: gap of unknown length
 67373 67452: contig of 80 bp in length
 67453 67552: gap of unknown length
 67553 72895: contig of 5343 bp in length
 72896 72995: gap of unknown length
 72996 79459: contig of 6474 bp in length
 79460 79569: gap of unknown length
 79570 87437: contig of 7868 bp in length
 87438 87537: gap of unknown length
 87538 95675: contig of 8138 bp in length
 95676 95775: gap of unknown length
 95776 103896: contig of 8121 bp in length
 103897 103966: gap of unknown length
 10397 112219: contig of 8223 bp in length
 112220 112319: gap of unknown length
 112320 126518: contig of 14199 bp in length
 126519 126618: gap of unknown length
 126619 139657: contig of 13039 bp in length
 139658 139757: gap of unknown length
 139758 155665: contig of 15908 bp in length
 155666 155765: gap of unknown length
 155766 173925: contig of 18160 bp in length
 173926 174025: gap of unknown length
 174026 213666: contig of 39641 bp in length.
 Location/Qualifiers
 1..213666="Mus musculus"
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="18"
 /map="18"
 /clone="RP23-59L5"
 /clone_lib="RP23 Female Mouse BAC"

```

misc_feature      1. .1333
                  /note="assembly_fragment"
gap               1334. .1433
                  /estimated_length=unknown
misc_feature      1434. .2451
                  /note="assembly_fragment"
gap               2452. .2551
                  /estimated_length=unknown
misc_feature      2552. .3569
                  /note="assembly_fragment"
gap               3570. .3669
                  /estimated_length=unknown
misc_feature      3670. .5895
                  /note="assembly_fragment"
gap               5896. .5995
                  /estimated_length=unknown
misc_feature      5996. .7002
                  /note="assembly_fragment"
gap               7003. .7102
                  /estimated_length=unknown
misc_feature      7103. .10154
                  /note="assembly_fragment"
gap               10155. .10254
                  /estimated_length=unknown
misc_feature      10255. .13202
                  /note="assembly_fragment"
gap               13203. .13302
                  /estimated_length=unknown
misc_feature      13303. .15342
                  /note="assembly_fragment"
gap               15343. .15442
                  /estimated_length=unknown
misc_feature      15443. .18266
                  /note="assembly_fragment"
gap               18267. .18366
                  /estimated_length=unknown
misc_feature      18367. .20817
                  /note="assembly_fragment"
gap               20818. .20917
                  /estimated_length=unknown
misc_feature      20918. .24056
                  /note="assembly_fragment"
gap               24057. .24156
                  /estimated_length=unknown
misc_feature      24157. .27612
                  /note="assembly_fragment"
gap               27613. .27712
                  /estimated_length=unknown
misc_feature      27713. .30777
                  /note="assembly_fragment"
gap               30778. .30877

Query Match      83.2%; Score 20.8; DB 14; Length 213666;
Best Local Similarity 91.7%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY               1 AAAAAAAAAAGCATGATGTGACA 24
Db               74529 AAAAAAAAAAGCATGATGTGACA 74552

RESULT 44
AC122327        217167 bp      DNA      linear      ROD 21-JUN-2005
LOCUS           sequence.
DEFINITION      Mus musculus BAC clone RP23-339D20 from chromosome 12, complete
ACCESSION       AC122327
VERSION         AC122327.5   GI:62868192
KEYWORDS        HTG.
SOURCE          Mus musculus (house mouse)
ORGANISM        Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

```

```

REFERENCE
AUTHORS         1 (bases 1 to 217167)
TITLE           Sciurognathi; Muridae; Murinae; Mus.
JOURNAL         Hodges, J., Bielicki, L. and Meyer, R.
                The sequence of Mus musculus BAC clone RP23-339D20
                Unpublished (2001)
AUTHORS         2 (bases 1 to 217167)
TITLE           McPherson, J.D. and Waterston, R.H.
JOURNAL         Direct Submission
                Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
                Parkway, St. Louis, MO 63108, USA
AUTHORS         3 (bases 1 to 217167)
TITLE           Wilson, R.K.
JOURNAL         Direct Submission
                Submitted (13-NOV-2004) Genome Sequencing Center, 4444 Forest Park
                Parkway, St. Louis, MO 63108, USA
AUTHORS         4 (bases 1 to 217167)
TITLE           Wilson, R.K.
JOURNAL         Direct Submission
                Submitted (23-APR-2005) Genome Sequencing Center, 4444 Forest Park
                Parkway, St. Louis, MO 63108, USA
AUTHORS         5 (bases 1 to 217167)
TITLE           Wilson, R.K.
JOURNAL         Direct Submission
                Submitted (21-JUN-2005) Genome Sequencing Center, Washington
                University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                MO 63108, USA
COMMENT         On Apr 23, 2005 this sequence version replaced gi:55734068.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BA033920

NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e. phred quality
>30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone, fosmid clone or direct clone walk sequence.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to obtain the consensus sequence. The
assembly was confirmed by restriction digest.
This finishing standard has slightly changed from the previous
Human standard. Specifically, standards for regions of low sequence
complexity (such as dinucleotide repeats and small unit tandem
repeats) have been relaxed. These regions are very prevalent in the
mouse genome, and the return on extended finishing efforts is
minimal.
If a sequence meets the criteria of the above statement, it needs
no comments or tags. If the criteria are not met, such as ambiguous
bases, then the region is duly annotated.

MAPPING INFORMATION:
This sequence was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The BAC library has been constructed by Kazuhiro Oseegawa and
Mitsuko Tateo in the laboratory of Pieter de Jong
(http://www.chori.org) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (http://www.tegen.com) or
Pieter de Jong and coworkers at http://www.chori.org

This sequence is the entire insert of the clone.
Location/Qualifiers

```


source

1. 217167
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="12"
/clone="RP23-339D20"
/clone_1db="RPC1-23"
109219..109236
/note="Sequence derived from one plasmid subclone."
193225..195100
/note="Unresolved tandem repeat."
216815..217162
/note="Unresolved simple sequence repeat."

Query Match 83.2% Score 20.8; DB 9; Length 217167;
Best Local Similarity 91.7%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGACA 24
Db 198139 AAAAAAAAAAGCATGATTGACA 198162

RESULT 45
AC109406
LOCUS
DEFINITION
AC109406 220189 bp DNA linear HTG 11-OCT-2002
Rattus norvegicus clone CH230-266P1, *** SEQUENCING IN PROGRESS
***, 11 unordered pieces.
AC109406 4 GI:23603940
HTG; HTG_PHASE1; HTG_DRAFT; HTG_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 220189)
Muzny,D,Marie, Metzker,M, Lee, S., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooke, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Bismalo, K., Blair, J., Blankenburg, K., Blych, P., Brown, W.,
Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J.,
Chickland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganter, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Kawth, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshuwa, L., Loulsegue, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, F.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackelmele, O., Okunolu, G., Olarunpusoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quito, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Swatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Woodden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Welts, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 220189)
Worley, K.C.
Direct Submission
Submitted (04-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220189)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:21737974.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: QDBA
Center clone name: CH230-266P1
----- Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 167000 bases at least Q40
Consensus quality: 170488 bases at least Q30
Consensus quality: 172720 bases at least Q20
Estimated insert size: 169991; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3064: contig of 3064 bp in length
* 3065 3164: gap of unknown length
* 3165 85762: contig of 82598 bp in length
* 85763 85862: gap of unknown length
* 85863 196631: contig of 110769 bp in length
* 196632 196731: gap of unknown length

```

* 196732 197842: contig of 1111 bp in length
* 197843 197942: gap of unknown length
* 197943 199179: contig of 1237 bp in length
* 199180 199279: gap of unknown length
* 199280 200964: contig of 1685 bp in length
* 200965 201064: gap of unknown length
* 201065 202874: contig of 1810 bp in length
* 202875 202974: gap of unknown length
* 202975 204490: contig of 1516 bp in length
* 204491 204590: gap of unknown length
* 204591 206200: contig of 1610 bp in length
* 206201 206300: gap of unknown length
* 206301 207513: contig of 1213 bp in length
* 207514 207614: gap of unknown length
* 207614 220189: contig of 12576 bp in length.
FEATURES
source
1. 220189
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-266P1"
3065..3164
/estimated_length=unknown
complement(6930..9830)
/misc_feature
/clone_end=Sp6
site:MboI
end_sequence:RXAEG85TV"
complement(80050..80921)
/misc_feature
/clone="clone_boundary"
clone_end:T7
site:MboI
end_sequence:RXAEG85TV"
85763..85862
/estimated_length=unknown
/misc_feature
/clone="wgs_end_extension"
clone_end:T7"
92405..94028
/clone="wgs_end_extension"
clone_end:T7"
196632..196731
/estimated_length=unknown
gap
197843..197942
/estimated_length=unknown
gap
199180..199279
/estimated_length=unknown
gap
200965..201064
/estimated_length=unknown
gap
202875..202974
/estimated_length=unknown
gap
204491..204590
/estimated_length=unknown
gap
206201..206300
/estimated_length=unknown
207514..207613
/estimated_length=unknown
gap
ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 220189;
Best Local Similarity 91.7%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAAGCATGATTGACA 24
Db 9253 AAAAAAAAAAAGCATGATTGACA 9276

```

```

VERSION
AC094468.6 GI:30466853
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 224364)
Muzny,D,Marle,M, Metzker,M, Lee, A, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Bunay, C., Burch, P., Cavazos, I., Ceasar, H., Center, A.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, Z., Chu, J.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Crease, A., D'Souza, L.,
Davila, M.L., Davis, C., Day-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshuwa, L., Louised, H., Lozado, R.U., Lu, X., Ma, U.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokelimeh, O., Okunou, G., Olarnpungoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C.,
Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Saverly, G., Scherer, S., Scott, G., Shetman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanal, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willison, R., Wlarczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, U., Zhou, J., Zhou, X., Zhou, S., Zhou, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.
Direct Submission
Unpublished
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 224364)
Worley, K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 224364)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24942366.
The sequence in this assembly is a combination of BAC based reads

```

and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GARZ

Center clone name: CH230-4A9

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 216693 bases at least Q40

Consensus quality: 218832 bases at least Q30

Consensus quality: 220333 bases at least Q20

Estimated insert size: 227307; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 224364: contig of 224364 bp in length.

Location/Qualifiers

1. 224364

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-4A9"

misc_feature

1. 1543

/note="wgs end-extension

clone end: 5p6"

2644..3492

/note="clone_boundary

clone_end: 5p6

site: EcorI

end_sequence: BH303785"

ORIGIN

Query Match

Best Local Similarity

Matches: 22; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 225581)
Direct Submission
Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 9, 2003 this sequence version replaced gi:31441851.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZK3124

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 217427 bases at least Q40

Consensus quality: 219130 bases at least Q30

Consensus quality: 220818 bases at least Q20

Insert size: 223881; sum-of-contigs

Insert size: 219000; 7.3% error; agarose-fp

Quality coverage: 6.82x in Q20 bases; sum-of-contigs Quality

coverage: 7.43x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently

* consists of 18 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 20602: contig of 20602 bp in length

* 20603 20702: gap of 100 bp

* 20703 57831: contig of 37129 bp in length

* 57832 57931: gap of 100 bp

* 57932 65164: contig of 7233 bp in length

* 65165 65264: gap of 100 bp

* 65265 71392: contig of 6128 bp in length

* 71393 71492: gap of 100 bp

* 71493 93677: contig of 22185 bp in length

* 93678 93777: gap of 100 bp

* 93778 101598: contig of 7821 bp in length

* 101599 101698: gap of 100 bp

* 101699 128689: contig of 26991 bp in length

* 128690 128789: gap of 100 bp

* 128790 138632: contig of 9843 bp in length

* 138633 138732: gap of 100 bp

* 138733 144193: contig of 5461 bp in length

* 144194 144293: gap of 100 bp

* 144294 155818: contig of 11525 bp in length

* 155819 155918: gap of 100 bp

* 155919 168354: contig of 12426 bp in length

* 168355 168454: gap of 100 bp

* 168455 171301: contig of 2847 bp in length

* 171302 171401: gap of 100 bp

* 171402 180364: contig of 8963 bp in length

* 180365 180464: gap of 100 bp

* 180465 188604: contig of 8140 bp in length

* 188605 188704: gap of 100 bp

* 188705 192855: contig of 4151 bp in length

* 192856 192955: gap of 100 bp

* 192956 201754: contig of 8799 bp in length

* 201755 201854: gap of 100 bp

* 201855 211372: contig of 9518 bp in length

* 211373 211472: gap of 100 bp

* 211473 225581: contig of 14109 bp in length.

FEATURES

source

1. 225581

Location/Qualifiers

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-3124"
/clone_1ib="DantolKey"
1..20602
/note="assembly fragment:01463
fragment_chain:1"
misc_feature
20703..57831
/note="assembly fragment:01892
fragment_chain:1"
misc_feature
57932..65164
/note="assembly fragment:02319
fragment_chain:1"
misc_feature
65265..71392
/note="assembly fragment:02166
fragment_chain:1"
misc_feature
71493..93677
/note="assembly fragment:01041
fragment_chain:1"
misc_feature
93778..101598
/note="assembly fragment:02235
fragment_chain:1"
misc_feature
101699..128689
/note="assembly fragment:01236
fragment_chain:1"
misc_feature
128790..138632
/note="assembly fragment:02758
fragment_chain:1"
misc_feature
138733..144193
/note="assembly fragment:02245
fragment_chain:1"
misc_feature
144294..155818
/note="assembly fragment:01852
fragment_chain:1"
misc_feature
155919..168354
/note="assembly fragment:01722
fragment_chain:1"
misc_feature
168455..171301
/note="assembly fragment:00772
fragment_chain:2"
misc_feature
171402..180364
/note="assembly fragment:02453
fragment_chain:2"
misc_feature
180465..188604
/note="assembly fragment:00524
fragment_chain:2"
misc_feature
188705..192855
/note="assembly fragment:02841
fragment_chain:2"
misc_feature
192956..201754
/note="assembly fragment:00644
fragment_chain:2"
misc_feature
201855..211372
/note="assembly fragment:02536
fragment_chain:2"
misc_feature
211473..225581
/note="assembly fragment:01373
fragment_chain:2"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14;
Best Local Similarity 91.7%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 2;

Length 225581;
Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24

DB 123690 AAAAAAAAAAGCATGAAAGTGACA 123667

RESULT 48
AC165350
LOCUS AC165350 231059 bp DNA linear HTG 10-AUG-2005

DEFINITION Mus musculus chromosome 12 clone RP23-249A17, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
ACCESSION AC165350
VERSION AC165350.2 GI:72098303
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 231059)

REFERENCE
AUTHORS
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
2 (bases 1 to 231059)

REFERENCE
AUTHORS
TITLE Direct Submision
JOURNAL Submitted (12-Jul-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 231059)

REFERENCE
AUTHORS
TITLE Direct Submision
JOURNAL Submitted (10-AUG-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Aug 10, 2005 this sequence version replaced gi:70721556.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0249A17

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 230013 bases at least Q40
Consensus quality: 230209 bases at least Q30
Consensus quality: 230496 bases at least Q20

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1497: contig of 1497 bp in length
* 1498 1597: gap of unknown length
* 1598 117894: contig of 116297 bp in length
* 117895 117994: gap of unknown length
* 117995 231059: contig of 113065 bp in length.

Location/Qualifiers
1..231059
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="12"
/clone="RP23-249A17"

FEATURES
source
1..1497
/note="assembly_name:Contig12"
1498..1597
/estimated_length=unknown
1598..117894
/note="assembly_name:Contig20
clone end:SP6
vector_side:right"
117895..117994

misc_feature
1..1497
/note="assembly_name:Contig12"
1498..1597
/estimated_length=unknown
1598..117894
/note="assembly_name:Contig20
clone end:SP6
vector_side:right"
117895..117994

misc_feature
1..1497
/note="assembly_name:Contig12"
1498..1597
/estimated_length=unknown
1598..117894
/note="assembly_name:Contig20
clone end:SP6
vector_side:right"
117895..117994

misc_feature
1..1497
/note="assembly_name:Contig12"
1498..1597
/estimated_length=unknown
1598..117894
/note="assembly_name:Contig20
clone end:SP6
vector_side:right"
117895..117994

misc_feature
1..1497
/note="assembly_name:Contig12"
1498..1597
/estimated_length=unknown
1598..117894
/note="assembly_name:Contig20
clone end:SP6
vector_side:right"
117895..117994

misc_feature
1..1497
/note="assembly_name:Contig12"
1498..1597
/estimated_length=unknown
1598..117894
/note="assembly_name:Contig20
clone end:SP6
vector_side:right"
117895..117994

misc_feature
1..1497
/note="assembly_name:Contig12"
1498..1597
/estimated_length=unknown
1598..117894
/note="assembly_name:Contig20
clone end:SP6
vector_side:right"
117895..117994

misc_feature 117895..231059
/note="assembly_name:Contig21
clone_end:T7
vector_side:left"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 231059;
Best Local Similarity 91.7%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 7905 AAAAAAAAAACCTGATTGTGACA 7928

RESULT 49
AE014836/c
LOCUS AE014836 250743 bp DNA linear INV 07-OCT-2002
DEFINITION Plasmodium falciparum 3D7 chromosome 11 section 1 of 8 of the
complete sequence.
ACCESSION AE014836 AE014186
VERSION AE014836.1 GI:23495940
KEYWORDS
SOURCE
ORGANISM Plasmodium falciparum 3D7
Plasmodium falciparum 3D7
Bukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 250743)
Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T.,
James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A.,
Kyes S., Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J.,
Angioli S., Petrea M., Allen J., Sengul J., Haft D.,
Mather M.W., Vaidya A.B., Martin D.M.A., Fairlamb A.H.,
Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I.,
Cummings L.M., Subramanian G.M., Mungall C., Venter J.C.,
Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Frazer C.M. and
Barrell B.
TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419 (6906), 498-511 (2002)
PUBMED 12368864
REFERENCE 2 (bases 1 to 250743)
Gardner M.J.
AUTHORS Direct Submission
JOURNAL Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
FEATURES
source
1..250743
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="11"
1..811
repeat_region /rpt_type=tandem
repeat_region 834..1482 /rpt_type=tandem
repeat_region 836..1443 /rpt_type=tandem
repeat_region 1168..1215 /rpt_type=tandem
repeat_region 1306..1338 /rpt_type=tandem
repeat_region 1709..1769 /rpt_type=tandem
repeat_region 2729..3310 /rpt_type=tandem
repeat_region 3524..4113 /rpt_type=tandem
repeat_region 4302..4428 /rpt_type=tandem
repeat_region 9249..9430 /rpt_type=tandem
repeat_region /rpt_type=tandem

repeat_region 9356..9430 /rpt_type=tandem
repeat_region 9673..9725 /rpt_type=tandem
repeat_region 9675..9725 /rpt_type=tandem
repeat_region 9724..9789 /rpt_type=tandem
repeat_region 9855..9897 /rpt_type=tandem
repeat_region 9939..10066 /rpt_type=tandem
repeat_region 9939..9990 /rpt_type=tandem
repeat_region 9999..10066 /rpt_type=tandem
repeat_region 10021..10149 /rpt_type=tandem
repeat_region 10036..10160 /rpt_type=tandem
repeat_region 10225..10814 /rpt_type=tandem
repeat_region 10231..10814 /rpt_type=tandem
repeat_region 10239..10568 /rpt_type=tandem
repeat_region 10652..10760 /rpt_type=tandem
repeat_region 10663..10814 /rpt_type=tandem
repeat_region 10666..10835 /rpt_type=tandem
repeat_region 10707..10814 /rpt_type=tandem
repeat_region 10882..10915 /rpt_type=tandem
repeat_region 10964..11046 /rpt_type=tandem
repeat_region 11208..11550 /rpt_type=tandem
repeat_region 11214..11288 /rpt_type=tandem
repeat_region 11324..11380 /rpt_type=tandem
repeat_region 11373..11462 /rpt_type=tandem
repeat_region 11386..11537 /rpt_type=tandem
repeat_region 11399..11534 /rpt_type=tandem
repeat_region 11662..12588 /rpt_type=tandem
repeat_region 12638..12670 /rpt_type=tandem
repeat_region 12653..16613 /rpt_type=tandem
repeat_region 13776..13848 /rpt_type=tandem
repeat_region 15068..18981 /rpt_type=tandem
repeat_region 17715..17867 /rpt_type=tandem
repeat_region 19143..19211 /rpt_type=tandem
repeat_region 19344..19445 /rpt_type=tandem
repeat_region 19346..19406 /rpt_type=tandem
repeat_region 19483..19753 /rpt_type=tandem
repeat_region 19742..19877 /rpt_type=tandem
repeat_region 20003..20227 /rpt_type=tandem

/rpt_type=tandem
repeat_region 20029..20089
/rpt_type=tandem
repeat_region 20128..20248
/rpt_type=tandem
repeat_region 20371..20435
/rpt_type=tandem
repeat_region 20472..20510
/rpt_type=tandem
repeat_region 20508..20680
/rpt_type=tandem
repeat_region 20508..20577
/rpt_type=tandem
repeat_region 20597..20749
/rpt_type=tandem
repeat_region 20895..21098
/rpt_type=tandem
repeat_region 20895..21077
/rpt_type=tandem
repeat_region 20931..21253
/rpt_type=tandem
repeat_region 20978..21099
/rpt_type=tandem
repeat_region 21010..21505
/rpt_type=tandem
repeat_region 21131..21405
/rpt_type=tandem
repeat_region 21608..21705
/rpt_type=tandem
repeat_region 21628..21881
/rpt_type=tandem
repeat_region 21628..21713
/rpt_type=tandem
repeat_region 21687..21878
/rpt_type=tandem
repeat_region 21693..21881
/rpt_type=tandem
repeat_region 22205..22240
/rpt_type=tandem
repeat_region 22325..22475
/rpt_type=tandem
repeat_region 22326..22475
/rpt_type=tandem
repeat_region 23932..24014
/rpt_type=tandem
/locus_tag="PF11_0007"
/locus_tag="PF11_0007"
/product="erythrocyte membrane protein 1 (pFEMP1)"
/protein_id="AAAS5604.1"
/db_xref="GI:23495941"
/translation="MEPHGSGGGGVDYIDHOSAKHLIDSIGKIIVHDOVEKEADYRSS
LQORLDATYSKPPDNOEKSPDCELTGHERTNATNTSTYPCRAKEERPSDYLGQC
TPOOIBGNDNRNGGACAPYRRLH.CNKNMWMMDNDNDSSKAKHLLIDVMAKYE
YSINTHYPKHOEKYKDTGTASOLCTVARSFADIGDIVRGRL.FRGNKSEKRRKE
KLEBNFKYTOQHEDVTSYSGKRRKALRPHKGDPTYOULEBDMNNRRKNRYAM
TEBAGSQYFRRTCSNGISBTNNKRCCTITETVPYFYVQYLHWEPMWEDCRKKK
HKLKDAIDKRGQVODADRYCDLNGYDCERTKGINMYRMDHCTGCFLSGSHRTWI
DKOEQFHKOKKIYDKITRGSDDHKGVEKKFYGKLKTDYKDVGKLELSDHEDVT
KPRDEGIIIDFKNNVSSSGGGGSGKTSHTYCOACPLCGVEKSGGNGNTWK
RKEDMPCPRLN.YKPIIDKGGITINFLVSGDREPTIGKTLN.FCLTHSGNSYVSSS
ARNGASRDKNKGSDDSDLYOKNTCYIDLEBKDEKEDGYDDPYDEVDYTGAGCLT
EKTNGEKRVNKQKTSHEMOKTFNPFYVWVHMLKDSIHRTKLLDKCINNNSSTAC
KNKKCKNKECECFQRMWITOKQOEWEQIKKAFRQKQINDVGGFFKLSHDVLQOVDK
DLILKSLOEAYGVEDIKHIDLEKEAALAAVAVLGGKNDVTTIDKLLNHEDKAKDCL
KKOECKEORGGVARSQTGPRPSAGDTVAHVEVEKEKEEKEEEDDEGDDGGDGG
DVQODNTVDVGAKEAPRVGSPPTTQNEVNPQIVNL.FSNPFGPNDACGLKXGPAP

Query Match 83.2%; Score 20.8; DB 2; Length 250743;
Best Local Similarity 91.7%; Pred. No. 2,3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 AAAAAAAAAAGCATGTTGTGACAC 25
Db 100565 AAAAAAAAAAGCATGTTGTGACAC 100542
RESULT 50
AC125986
LOCUS
DEFINITION
Rattus norvegicus clone CH230-3K20, WORKING DRAFT SEQUENCE.
AC125986
AC125986.4 GI:30581569
VERSION
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE
Rattus norvegicus
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurgnathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 263479)
REFERENCE
Muzny,D,Marie,,Metzker,M,Leee,,Abramson,S,,Adams,C,,Alder,J,,
Allen,C,,Allen,H,,Alsbrooks,S,,Amit,A,,Anguiano,D,,
Anyalebechi,V,,Aoyagi,A,,Ayodeji,M,,Baca,E,,Baden,H,,
Baldwin,D,,Bandaranaika,D,,Barber,M,,Barnstead,M,,Benhamed,F,,
Biswal,K,,Blair,J,,Blankenburg,K,,Blyth,P,,Brown,M,,
Bryant,N,,Bunay,C,,Burch,P,,Butrell,K,,Calderton,E,,
Cardenas,V,,Carter,K,,Cavazos,I,,Cesar,H,,Center,A,,
Chacko,J,,Chavez,D,,Chen,G,,Chen,R,,Chen,Y,,Chen,Z,,Chu,J,,
Cleveland,C,,Cockrell,R,,Cox,C,,Coyle,M,,Crete,A,,D'Souza,L,,
Davila,M,L,,Davis,C,,Davy-Carroll,L,,De Anda,C,,Dederich,D,,
Delgado,O,,Denson,S,,Deramo,C,,Ding,Y,,Dinh,H,,Diva,K,,
Draper,H,,Dugan-Rocha,S,,Dunn,A,,Durbin,K,,Duval,B,,Eaves,K,,
Egan,A,,Escotto,M,,Eugene,C,,Evans,C,A,,Falls,T,,Fan,G,,
Fernandez,S,,Finley,M,,Flagg,N,,Forbes,L,,Foster,M,,Foster,P,,
Frazer,C,M,,Gabisi,A,,Ganta,R,,Garcia,A,,Garner,T,,Garza,M,,
Gebregiorgis,E,,Geer,K,,Gilli,R,,Grady,M,,Guerra,M,,Guevara,W,,
Gunaratne,P,,Haaland,W,,Hamill,C,,Hamilton,K,,
Harvey,Y,,Havlak,P,,Hawes,A,,Henderson,N,,Hernandez,J,,
Hernandez,R,,Hines,S,,Hladun,S,L,,Hodgson,A,,Hogues,M,,
Hollins,B,,Howells,S,,Huljk,S,,Hume,J,,Idlebird,D,,Jackson,A,,
Jackson,L,,Jacob,L,,Jiang,H,,Johnson,B,,Johnson,R,,Jolivet,A,,
Karpatny,S,,Kelly,S,,Kelly,S,,Khan,Z,,King,L,,Kovar,C,,
Kowis,C,,Kraft,C,L,,Lebow,H,,Levan,J,,Lewis,L,,Li,Z,,Liu,J,,
Liu,J,,Liu,W,,Liu,Y,,London,P,,Longacre,S,,Lopez,J,,
Lorensuewa,L,,Louieged,H,,Lozado,R,J,,Lu,X,,Ma,J,,
Maheshwari,M,,Mahindaratne,M,,Mahmoud,M,,Mallory,K,,Mangum,A,,
Mangum,B,,Mapua,P,,Martin,K,,Martin,R,,Martinez,E,,
Mawhney,S,,McLeod,M,P,,McNeill,T,Z,,Meenen,E,,
Milosavljevic,A,,Miner,G,,Minja,E,,Montemayor,J,,Moore,S,,
Morgan,M,,Morris,K,,Morris,S,,Mundassa,M,,Murphy,M,,Nair,L,,
Nankervis,C,,Neal,D,,Newton,N,,Nguyen,N,,Norris,S,,
Nwackemele,O,,Okunou,G,,Olarinmugoon,A,,Pal,S,,Parks,K,,
Pasternak,S,,Paul,H,,Perez,A,,Perez,L,,Pfeankoch,C,,
Plopper,F,,Pondexter,A,,Popovic,D,,Primus,E,,Pu,L,,L,,
Puzo,M,,Quitoz,J,,Rachlin,E,,Reeves,K,,Regier,M,A,,Reigh,R,,
Reilly,B,,Reilly,M,,Ren,Y,,Reuter,M,,Richards,S,,Riggs,F,,
Rivers,C,,Rodkey,T,,Rojas,A,,Rose,M,,Rose,R,,Ruiz,S,J,,
Sanders,W,,Savery,G,,Scherer,S,,Scott,G,,Shatsman,S,,Shen,H,,
Shetty,J,,Shvartsbeyn,A,,Sison,I,,Sitter,C,D,,Smajls,D,,
Sneed,A,,Sodergran,B,,Song,X,-Z,,Soreller,R,,Soes,J,,
Steinle,M,,Strong,R,,Sutton,A,,Swatek,A,,Tabor,P,,Taylor,C,,
Taylor,T,,Thomas,N,,Thomas,S,,Tingay,A,,Tresos,Z,,Umanu,K,,
Valas,R,,Vera,V,,Villaana,D,,Waldron,L,,Walker,B,,Wang,J,,
Wang,Q,,Wang,S,,Warren,J,,Warren,R,,Wleczky,R,,Wooden,H,,Mortley,K,,
Williams,G,,Willison,R,,Wleczky,R,,Wooden,H,,Mortley,K,,
Wright,D,,Wright,K,,Wu,J,,Yakub,S,,Yen,J,,Yoon,L,,Yoon,V,,
Yu,F,,Zhang,J,,Zhou,J,,Zhou,X,,Zhou,S,,Zhou,D,,von
Miednerhausen,A,,Weiss,R,,Smith,D,R,,Holt,R,A,,Smith,H,O,,
Weinstock and Gibbs,R,A.
TITLE
JOURNAL
Unpublished

REFERENCE 2 (bases 1 to 263479)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 3 (bases 1 to 263479)
 AUTHORS Rat Genome Sequencing Consortium.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GKN
 Center clone name: CH230-3K20
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 254594 bases at least Q40
 Consensus quality: 256593 bases at least Q30
 Consensus quality: 257885 bases at least Q20
 Estimated insert size: 269020; sum-of-contigs estimation
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 1 263479: contig of 263479 bp in length.

FEATURES
 source
 1..263479
 location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-3K20"

misc_feature
 1..1524
 /note="wgs_end_extension
 clone_end:T7"

misc_feature
 2251..2443
 /note="clone_boundary
 clone_end:T7
 site:Ecori

misc_feature
 complement(244394..258856)
 /note="clone_boundary
 clone_end:Sp6
 site:Ecori
 end_sequence:BH312405"

misc_feature 259735..261171
 /note="wgs_end_extension
 clone_end:Sp6"
 misc_feature 261222..263479
 /note="wgs_end_extension
 clone_end:Sp6"

Query Match 83.2%; Score 20.8; DB 14; Length 263479;
 Best Local Similarity 91.7%; Pred. No. 2.3e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACATGATTGTGACA 24
 |||||
 Db 216679 AAAAAAAAAACCATGATTGTGACA 216702

Search completed: December 14, 2005, 11:11:16
 Job time : 875.8 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 / Search time 203.2 Seconds
(without alignments) 819.967 Million cell updates/sec

Title: US-10-681-773-6

Perfect score: 25
Sequence: 1 aaaaaaaaaagcattgtgtgacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database:

1: N_Geneseq_21:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000s:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	84.0	32404	9	ADA02894 Human BLR
2	21	84.0	32404	10	ADB72632 Human BLR
3	21	84.0	32404	10	ADC85373 Mouse BLR
4	21	84.0	32404	12	ADM74489 Human car
5	21	84.0	32404	13	ADR66970 Human can
6	21	84.0	33296	14	AD212715 Human can
7	20.8	83.2	2013	3	AAC42808 Human can
8	20.8	83.2	2013	6	ABZ13714 Arabidops
9	20.2	80.8	25032	9	ADA02495 Mouse Myc
10	20.2	80.8	25032	10	ADB72233 Mouse Myc
11	20.2	80.8	25032	10	ADB82935 Mouse Myc
12	20.2	80.8	25032	10	ADB82935 Mouse Myc
13	20.2	80.8	16052	4	AAD26297 Human gly
14	19.8	79.2	110000	11	ACN43964_2 Human gly
15	19.8	79.2	110000	11	ACN43964_3 Human gly
16	19.8	79.2	177380	8	ACF62751 Cancer ba
17	19.8	79.2	177380	8	ADB20870 MRP1 base
18	19.8	79.2	177380	10	ADB87959 Human UGT
19	19.8	79.2	177380	10	ADB96942 Human MDR

20	19.8	79.2	177380	10	ADB92133 Human MDR
21	19.4	77.6	329	6	ABL1814 Human ova
22	19.4	77.6	356	6	ABL64629 Stomach c
23	19.4	77.6	356	6	ABL60674 Human ova
24	19.4	77.6	377	6	ABL87666 Human ova
25	19.4	77.6	393	6	ABLA2013 Human ova
26	19.4	77.6	396	6	ABL7751 Human ova
27	19.4	77.6	553	4	AAH11468 Human ova
28	19.4	77.6	3434	4	AAH14440 Human CDN
29	19.4	77.6	3434	13	ADV35067 Human CDN
30	19.4	77.6	219352	13	ABD33098 Human CDN
31	19.2	76.8	163	13	ACN48867 Human pr
32	19.2	76.8	166	13	ACN48946 Human pr
33	19.2	76.8	414	4	AAU00715 Human rep
34	19.2	76.8	464	4	AAH14617 Human rep
35	19.2	76.8	464	4	ABA56346 Human foe
36	19.2	76.8	464	4	AAI35989 Human foe
37	19.2	76.8	464	4	ABA45831 Human bre
38	19.2	76.8	464	4	ABA25987 Human bre
39	19.2	76.8	464	4	AAK30026 Human bon
40	19.2	76.8	464	4	AAK04525 Human bra
41	19.2	76.8	464	4	ABS29678 Human liv
42	19.2	76.8	464	6	ABS04598 Human gen
43	19.2	76.8	516	12	ACH77614 Human gen
44	19.2	76.8	568	13	ACN48254 Cotton pr
45	19.2	76.8	782	3	AAZ97283 Human pro
46	19.2	76.8	1654	13	ADT17668 Human CDN
47	19.2	76.8	1928	8	ADA71787 Human CDN
48	19.2	76.8	2000	8	ADA71787 Rice gene
49	19.2	76.8	2366	5	AAE77275 Human alp
50	19.2	76.8	8050	3	AAE52156 A. thalia
51	19.2	76.8	30360	4	AAK82759 Human imm
52	19.2	76.8	34267	13	ADV87725 Streptoco
53	19.2	76.8	34267	13	ADV78978 Streptoco
54	19.2	76.8	39887	4	AAK81263 Human imm
55	19.2	76.8	39887	4	AAK79153 Human imm
56	19.2	76.8	110000	6	ABN71527_03 Continuation (4 of
57	19.2	76.8	110000	13	ADV81204_04 Continuation (5 of
58	19.2	76.8	110000	14	AEA61163_3 Continuation (4 of
59	19.2	76.8	110000	14	AEA61163_4 Continuation (5 of
60	19.2	76.8	188267	14	AD213837 Human can
61	19.2	76.8	188267	12	AD059476 Human can
62	18.8	75.2	133	3	AACT17912 Human sec
63	18.8	75.2	550	9	ACH43937 Human foe
64	18.8	75.2	845	4	AAK63853 Human imm
65	18.8	75.2	3246	2	AAK89525 Polynucle
66	18.8	75.2	3246	14	AD249659 Insulin b
67	18.8	75.2	3246	14	ABE22785 Human col
68	18.8	75.2	3325	4	AAK6073 Human imm
69	18.8	75.2	3912	12	ADU12722 Human neu
70	18.8	75.2	13161	5	AAE97867 Human gen
71	18.8	75.2	25619	4	AAE26966 Human gen
72	18.8	75.2	25619	5	ABK72116 Human rep
73	18.8	75.2	25619	5	ABK72116 Human ova
74	18.8	75.2	25619	6	ABK91708 Human ger
75	18.8	75.2	32522	12	ADU12442 DNA fragm
76	18.8	75.2	51705	10	ADU10262 Human ger
77	18.8	75.2	51705	9	ADA02723 Mouse Syk
78	18.8	75.2	51705	10	ADB72461 Mouse Syk
79	18.8	75.2	51705	10	ADB95971 Mouse Syk
80	18.6	74.4	519	6	ABX31946 Plant ful
81	18.6	74.4	594	6	ABX53634 Human eos
82	18.6	74.4	597	4	AAI18280 Probe #82
83	18.6	74.4	597	4	ABA63264 Human foe
84	18.6	74.4	597	4	AAI43336 Probe #12
85	18.6	74.4	597	4	ABA30494 Probe #89
86	18.6	74.4	597	4	AAK37474 Human bon
87	18.6	74.4	597	4	AAK11731 Human bra
88	18.6	74.4	597	4	ABK37154 Human liv
89	18.6	74.4	597	6	ABK11453 Human gen
90	18.6	74.4	642	3	AAE32609 Arabidops
91	18.6	74.4	650	6	ABO65950 Arabidops
92	18.6	74.4	761	12	AD062623 Transcrip

```
93 18.6 74.4 860 11 ACN83819 Acn83819 Breast ca
94 18.6 74.4 1612 10 ADE61980 Ade61980 Rat gene
95 18.6 74.4 1612 10 ADE61984 Ade61984 Rat gene
96 18.6 74.4 2000 8 ADA73383 Ada73383 Rice gene
97 18.6 74.4 2517 6 AAL11167 Aal11167 Wooden le
98 18.6 74.4 2688 6 AAL11168 Aal11168 Wooden le
99 18.6 74.4 3226 13 ADV97351 Adv97351 N tabacum
100 18.6 74.4 4001 6 AAL11166 Aal11166 Wooden le
101 18.6 74.4 4860 4 AAK69520 Aak69520 Human imm
102 18.6 74.4 5020 4 AAH18707 Aah18707 Human imm
103 18.6 74.4 5080 4 ABK42311 Abk42311 Genomic s
104 18.6 74.4 5080 4 ABK42310 Abk42310 Genomic s
105 18.6 74.4 5080 9 ADB60466 Adb60466 Connectiv
106 18.6 74.4 5080 9 ADB60467 Adb60467 Connectiv
107 18.6 74.4 6495 5 AAL06441 Aal06441 Human rep
108 18.6 74.4 6495 5 AAS40553 Aas40553 DNA encod
109 18.6 74.4 6495 11 ADJ09759 Adj09759 Human pro
110 18.6 74.4 7028 6 AAL11165 Aal11165 Wooden le
111 18.6 74.4 7450 4 ABL25450 Abl25450 Drosophi1
112 18.6 74.4 14779 4 ABL27058 Abl27058 Drosophi1
113 18.6 74.4 37698 11 ACN43854 Acn43854 Human gen
114 18.6 74.4 68255 11 ACN44362 Aal41152 Wooden le
115 18.6 74.4 70768 6 AAL11152 Aal11152 Wooden le
116 18.6 74.4 91617 14 ADZ13161 Aa10225 Human PCT
117 18.6 74.4 106746 3 AAB10225 Aa10225 Human PCT
118 18.6 74.4 110000 13 ABD32594_3 Aac65548 Human kin
119 18.6 74.4 121162 3 AAC65548 Aac65548 Human kin
120 18.6 74.4 130207 11 ACN44762 Acn44762 Human gen
121 18.6 74.4 138169 3 AAA34791 Aaa34791 Human ade
122 18.6 74.4 138837 13 ABD33163 Abd33163 Human can
123 18.6 74.4 141586 11 ABD20685 Abd20685 Human pul
124 18.6 74.4 141589 3 AAA35005 Aaa35005 Human ade
125 18.6 74.4 141589 3 AAA35030 Aaa35030 Human ade
126 18.6 74.4 141589 3 AAF21152 Aaf21152 Human ELA
127 18.6 74.4 141589 3 AAF20913 Aaf20913 Human ELA
128 18.6 74.4 141589 3 AAF21127 Aaf21127 Human low
129 18.6 74.4 141589 10 ABZ96821 Abz96821 Human nuc
130 18.6 74.4 141589 10 ABZ96807 Abz96807 Human ELA
131 18.6 74.4 141589 10 ABZ96846 Abz96846 Human nuc
132 18.6 74.4 141589 11 ABD20670 Abd20670 Human pul
133 18.6 74.4 141601 11 ABD19162 Abd19162 Human ELA
134 18.6 74.4 145831 6 ABL69213 Abl69213 Prostate
135 18.6 74.4 145831 6 ABL68806 Abl68806 Lung canc
136 18.6 74.4 145831 6 ABL68588 Abl68588 Kidney ca
137 18.6 74.4 145831 6 ABL62309 Abl62309 Colon ade
138 18.6 74.4 145831 6 ABT10149 Abt10149 Human bre
139 18.6 74.4 146981 3 AAF21442 Aaf21442 Human ELA
140 18.6 74.4 146982 10 ABZ97136 Abz97136 Human ELA
141 18.6 74.4 146984 11 ABD19160 Abd19160 Human ELA
142 18.6 74.4 154875 14 AEA61197 Aea61197 Human MTH
143 18.6 74.4 163321 11 ACN43898 Acn43898 Human gen
144 18.6 74.4 175077 11 ACN44626 Aacn44626 Human gen
145 18.6 74.4 209273 3 AAF21437 Aaf21437 Human fac
146 18.6 74.4 209274 10 ABZ97131 Abz97131 Human enz
147 18.6 74.4 209284 11 ABD17970 Abd17970 Human fac
148 18.4 73.6 35651 4 AAF57595 Aaf57595 ATM compl
149 18.4 73.6 150351 13 ABD33360 Abd33360 Murine ca
150 18.2 72.8 388 13 ACF86423 Acf86423 Human SIR
```

ALIGNMENTS

```
RESULT 1
ADA02894
ID ADA02894 standard; DNA; 32404 BP.
```

```
AC ADA02894;
XX
XX 06-NOV-2003 (first entry)
XX Human BLR1 carcinoma associated gene, SEQ ID NO:1412.
XX
```

```
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
OS Homo sapiens.
PN WO2003057146-A2.
XX
XX 17-JUL-2003.
PD
PD 17-JUL-2003.
PF 26-DEC-2002; 2002WO-US041414.
XX
XX 26-DEC-2001; 2001US-00035832.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
XX WPI; 2003-587068/55.
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1412; 245bp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
XX acid sequences from mouse and human (ADA01482-ADA03094), and to
XX recombinant carcinoma associated proteins (CAP) encoded by them. The
XX invention also encompasses expression vectors and host cells comprising a
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically
XX binds to the protein, and a biochip comprising CA nucleic acid or
XX fragments thereof. The sequences of the invention were identified using
XX oncogenic retroviruses, which insert into the genome of the host organism
XX at random. Many of these do not carry transduced host oncogenes or
XX pathogenic trans-acting viral genes, meaning that cancer incidence is a
XX direct consequence of the effects of proviral integration into host
XX protooncogenes. The CA nucleic acid sequences can be used to diagnose
XX carcinoma (especially breast cancer, prostate cancer, lymphoma or
XX leukemia) or a propensity to carcinoma by determination of the sequence
XX of a CA gene, or by determination of CA gene expression in particular
XX tissues. CA nucleic acid, proteins and antibodies are also useful as
XX therapeutic agents and in screening and evaluating drug candidates. The
XX present sequence represents a specifically claimed human CA nucleic acid
XX sequence of the invention. Note: The complete sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;
XX
XX Query Match 84.0%; Score 21; DB 9; Length 32404;
XX Best Local Similarity 100.0%; Pred. No. 1.2e+02;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAGCATGATTGTG 21
XX |||||||
XX 302 AAAAAAAAAAGCATGATTGTG 322
XX
RESULT 2
ADB72632
ID ADB72632 standard; DNA; 32404 BP.
XX
XX ADB72632;
XX
XX 04-DEC-2003 (first entry)
XX Human BLR1 gene.
XX
XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX Homo sapiens.
OS
```

XX WO2003008583-A2.
 PN
 XX
 PD 30-JAN-2003.
 XX
 PF 26-DEC-2001; 2001WO-US051291.
 XX
 PR 02-MAR-2001; 2001US-00798586.
 PR 23-OCT-2001; 2001US-00004113.
 PR 08-NOV-2001; 2001US-00052482.
 PR 30-NOV-2001; 2001US-00997722.
 PR 20-DEC-2001; 2001US-00034650.
 XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW, Engelhard EK;
 XX
 DR WPI; 2003-239337/23.
 XX
 PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.
 XX
 PS Claim 1; SEQ ID NO 460; 2304bp; English.
 XX
 CC The invention relates to a novel recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the 660 sequences fully defined
 CC in the specification. A polynucleotide of the invention has cytosstatic
 CC activity, and may have a use in gene therapy, or in a vaccine. The
 CC recombinant nucleic acids and polypeptides are useful for treating
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC sarcomas. The present sequence represents a human gene of the invention.
 XX
 SQ Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;
 XX
 QY Query Match 84.0%; Score 21; DB 10; Length 32404;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTG 21
 |||||
 302 AAAAAAAAAAGCATGATTGTG 322
 Db
 RESULT 3
 ADC85373
 ID ADC85373 standard; DNA; 32404 BP.
 XX
 AC ADC85373;
 XX
 DT 01-JAN-2004 (first entry)
 XX
 DE Mouse B1rl coding sequence.
 XX
 KM Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
 KM secreted; transmembrane; intracellular; ds.
 XX
 OS Mus sp.
 XX
 PN WO2003045230-A2.
 XX
 PD 05-JUN-2003.
 XX
 PF 02-DEC-2002; 2002WO-US038582.
 XX
 PR 30-NOV-2001; 2001US-00997722.
 XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW, Engelhard EK;
 XX
 DR WPI; 2003-513603/48.
 XX
 PT New recombinant nucleic acid comprising a nucleotide sequence of any of

PT the carcinoma-associated (CA) genes, useful for screening for drug
 PT candidates for diagnosing or treating carcinomas.
 XX
 PS Claim 1; SEQ ID NO 159; 983bp; English.
 XX
 CC The invention relates to a recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the fully defined carcinoma-
 CC associated (CA) genes from the 50 tables given in the specification. The
 CC CA proteins are secreted, transmembrane or intracellular proteins. The
 CC recombinant nucleic acids are useful for screening for drug candidates
 CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
 CC ADC85314 represent CA genes of the invention.
 XX
 SQ Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other;
 XX
 QY Query Match 84.0%; Score 21; DB 10; Length 32404;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTG 21
 |||||
 302 AAAAAAAAAAGCATGATTGTG 322
 Db
 RESULT 4
 ADM74489
 ID ADM74489 standard; DNA; 32404 BP.
 XX
 AC ADM74489;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Human carcinoma associated (CA) nucleic acid #79.
 XX
 KM Human, carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
 KM carcinoma associated protein; CAP; carcinoma; leukemia; lymphoma;
 KM cytosstatic.
 XX
 OS Homo sapiens.
 XX
 PN US2004072154-A1.
 XX
 PD 15-APR-2004.
 XX
 PF 30-NOV-2001; 2001US-00997722.
 XX
 PR 22-DEC-2000; 2000US-00747377.
 PR 02-MAR-2001; 2001US-00798586.
 XX
 PA (MORR/) MORRIS D W.
 PA (ENGE/) ENGELHARD E K.
 XX
 PI Morris DW, Engelhard EK;
 XX
 DR WPI; 2004-328562/30.
 XX
 PT New carcinoma associated gene or protein, useful for preparing a
 PT composition for diagnosing or treating carcinoma e.g., leukemia or
 PT lymphoma.
 XX
 PS Claim 1; SEQ ID NO 160; 29pp; English.
 XX
 CC The invention relates to new recombinant nucleic acids. The invention
 CC also relates to a host cell comprising a recombinant nucleic acid or
 CC expression vector, an expression vector comprising a recombinant nucleic
 CC acid, a recombinant protein, a method of screening for drug candidates, a
 CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
 CC method of screening for a bioactive agent capable of modulating the
 CC activity of a CAP, a method of evaluating the effect of a candidate
 CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
 CC the activity of a CAP, a method of treating carcinomas, a method of
 CC neutralizing the effect of a CAP and a method of diagnosing carcinoma or

propensity to carcinoma. A method of evaluating the effect of a candidate carcinoma drug comprises administering the drug to a patient, removing a cell sample from the patient and determining alterations in the expression or activation of a gene comprising the nucleotide sequence. A method of diagnosing carcinoma comprises determining the expression of one or more genes comprising the nucleic acid sequence in a first tissue type of a first individual and comparing the expression of the gene from a second normal tissue type from the first individual or a second unaffected individual, where a difference in the expression indicates that the first individual has carcinoma. A method of inhibiting the activity of a CAP comprises binding an inhibitor to the CAP. Treating carcinomas comprises administering to a patient an inhibitor of CAP. Neutralising the effect of a CAP comprises contacting an agent specific for the CAP. The polypeptide specifically binds to the protein encoded by the nucleic acid. It comprises an antibody that specifically binds to the protein encoded by the nucleic acid. The nucleic acids are useful for preparing a composition for diagnosing or treating carcinoma e.g., leukemia or lymphoma. This sequence represents a human carcinoma associated (CA) nucleic acid of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.

Seq Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other:

Query Match 84.0%; Score 21; DB 13; Length 32404;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 302 AAAAAAAAAAGCATGATTGTG 322

RESULT 5
AD212715
ID AD212715 standard; DNA; 32404 BP.
AC ADR66970;
XX ADR66970;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer associated gene genomic sequence SEQ ID NO:16.
XX
KW cancer; cancer associated nucleic acid; cancer associated gene;
KW cancer associated protein; CAP; cytostatic; vaccine; gene therapy;
KW lymphoma; leukaemia; human; gene; ds.
XX
OS Homo sapiens.
XX
XX WO2004074321-A2.
XX
XX
XX 02-SEP-2004.
XX
XX
XX 17-FEB-2004; 2004WO-US005000.
XX
XX
XX 14-FEB-2003; 2003US-00367094.
XX
XX 14-MAR-2003; 2003US-0038838.
XX
XX 23-SEP-2003; 2003US-00669920.
XX
XX 15-DEC-2003; 2003US-00737318.
XX
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX
XX Morris DW, Malandro MS;
XX
XX
XX MPI; 2004-652915/63.
XX
XX P-PSDB; ADR66972.
XX
XX
XX New isolated cancer-associated polynucleotides and polypeptides useful
XX PT for diagnosing, preventing or treating cancers, especially lymphoma and
XX PT leukemia, or in screening for agents that modulate cancer.
XX
XX Claim 16; SEQ ID NO 16; 166bp; English.

The present invention describes an isolated cancer associated (CA) nucleic acid (1). Also described: (1) an expression vector comprising (1) a host cell comprising (1) or the expression vector; (3) a microarray for detecting a CA nucleic acid; (4) an isolated cancer associated protein (CAP) polypeptide, encoded within an open reading frame of a CA sequence; (5) an isolated antibody, or its antigen binding fragment, that binds to the above polypeptide; (6) a hybridoma that produces the above monoclonal antibody; (7) a pharmaceutical composition comprising the above antibody and a pharmaceutical excipient; (8) a kit for detecting cancer cells, comprising the (monoclonal) antibody described above; (9) methods for diagnosing cancer or for detecting the presence or absence of cancer cells in an individual; (10) a method for inhibiting growth of cancer cells in an individual; (11) a method for delivering a therapeutic agent to cancer cells in an individual; (12) an electronic library comprising the above polynucleotide or polypeptide, or their fragments; (13) methods of screening for anticancer activity or for a bioactive agent capable of modulating the activity of a CAP; (14) methods for detecting cancer associated with expression of a polypeptide in a test cell sample, or with the presence of an antibody in a test serum sample; (15) a method for treating cancer; and (16) a method for inhibiting the expression of CA gene in a cell. The CA sequences have cytostatic activity, and can be used in vaccines, and in gene therapy. The composition and methods are useful for detecting, diagnosing, preventing and treating cancers, especially lymphoma and leukaemia. They may also be used in screening for agents that modulate cancer. The present sequence represents a cancer associated gene genomic DNA sequence, which is used in the exemplification of the present invention.

Seq Sequence 32404 BP; 7711 A; 8273 C; 9097 G; 7303 T; 0 U; 20 Other:

Query Match 84.0%; Score 21; DB 13; Length 32404;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 302 AAAAAAAAAAGCATGATTGTG 322

RESULT 6
AD212715
ID AD212715 standard; DNA; 33296 BP.
AC AD212715;
XX
XX
DT 16-JUN-2005 (first entry)
XX
DE Human cancer-associated genomic DNA #19.
XX
XX
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasia;
KW cytostatic; gene; ds.
XX
XX
XX Homo sapiens.
XX
XX
XX WO2005031001-A2.
XX
XX
XX 07-APR-2005.
XX
XX
XX 23-SEP-2004; 2004WO-US031617.
XX
XX
XX 23-SEP-2003; 2003US-00669920.
XX
XX
XX (CHIR) CHIRON CORP.
XX
XX
XX Morris DW, Malandro MS;
XX
XX
XX MPI; 2005-273395/28.
XX
XX
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
XX PT comprises two or more nucleic acid probes.
XX
XX
XX Disclosure; SEQ ID NO 235; 199bp; English.

XX The invention relates to a nucleic acid array for detecting a cancer
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC The invention also relates to a peptide array comprising two or more
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC that binds to a polypeptide, an isolated antibody or its fragment which
CC binds to a polypeptide, which is prepared by immunizing a host animal
CC with a composition comprising the polypeptide or its antigen binding
CC fragment and collecting cells from the host expressing antibodies against
CC the antigen or its antigen binding fragment, a composition comprising the
CC antibody and a carrier, a method of screening for anticancer activity, a
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC method of treating cancer and a method of inhibiting expression of a CA
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC nucleic acids. The antibody is useful for detecting the presence or
CC absence of cancer cells in an individual which involves contacting cells
CC from the individual with the antibody and detecting a complex of a CA
CC protein from the cancer cells and the antibody, where the detection of
CC the complex correlates with the presence of cancer cells in the
CC individual. The composition is useful for inhibiting growth of cancer
CC cells in an individual or for delivering a therapeutic agent to cancer
CC cells in an individual. The invention is also useful for diagnosing
CC cancer, for treating cancer and for inhibiting expression of a CA gene in
CC a cell. This sequence represents human cancer-associated genomic DNA of
CC the invention.
XX
SQ Sequence 33296 BP, 7974 A, 8525 C, 9288 G, 7509 T, 0 U, 0 Other;
Query Match 84.0%; Score 21; DB 14; Length 33296;
Best Local Similarity 100.0%; Pred. No. 1,2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 676 AAAAAAAAAAGCATGATTGTG 696
RESULT 7
AAC2808/C
ID AAC2808 standard; DNA; 2013 BP.
XX
AC AAC2808;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 36926.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135533P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.

```

PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147322P.
PR 05-AUG-1999; 99US-0147326P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154079P.
PR 20-SEP-1999; 99US-0154739P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158212P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159283P.
PR 13-OCT-1999; 99US-0159284P.
PR 13-OCT-1999; 99US-0159285P.
PR 14-OCT-1999; 99US-0159310P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.

```

```

PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      83.2%; Score 20.8; DB 3; Length 2013;
Best Local Similarity 91.7%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 1329 AAAAAACAAAGCATGACTGTGACA 1306

RESULT 8
ABZ13714/C
ID ABZ13714 standard; DNA; 2013 BP.
XX
AC ABZ13714;
XX
DT 21-JUN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1519.
XX
KM Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
PD 28-FEB-2002.
PF 24-AUG-2001; 2001WO-US026685.
XX
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI ) SCRIPTS RES INST.
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
XX producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 14; SEQ ID NO 1519; 577bp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants

```

CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
 CC in methods of the invention. Note: The sequence data for this patent is
 CC not represented in the printed specification, but is based on sequence
 CC information supplied to Derwent by the European Patent Office
 CC
 SQ Sequence 2013 BP, 545 A, 384 C, 490 G, 594 T, 0 U, 0 Other;

Query Match 83.2%; Score 20.8; DB 6; Length 2013;
 Best Local Similarity 91.7%; Pred. No. 1.2e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGACA 24
 Db 1329 AAAAAAAAAAGCATGACTGTGACA 1306

RESULT 9
 ADA02495/C
 ID ADA02495 standard; DNA; 25032 BP.

ADA02495;

06-NOV-2003 (first entry)

Mouse Myc carcinoma associated gene, SEQ ID NO:1013.

Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
 prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
 gene; ds.

Mus sp.

MO2003057146-A2.

17-JUL-2003.

26-DEC-2002; 2002WO-US041414.

26-DEC-2001; 2001US-00035832.

(SAGR-) SAGRES DISCOVERY.

Morris DW;

WPI; 2003-587068/55.

New recombinant nucleic acid encoding carcinoma associated protein,
 useful for preparing compositions for treating carcinomas.

Claim 1; SEQ ID NO 1013; 245pp; English.

The invention relates to recombinant carcinoma associated (CA) nucleic
 acid sequences from mouse and human (ADA01482-ADA03094), and to
 recombinant carcinoma associated proteins (CAP) encoded by them. The
 invention also encompasses expression vectors and host cells comprising a
 CA nucleic acid, a polypeptide (especially an antibody) that specifically
 binds to the protein, and a biochip comprising CA nucleic acid or
 fragments thereof. The sequences of the invention were identified using
 oncogenic retroviruses, which insert into the genome of the host organism
 at random. Many of these do not carry transduced host oncogenes or
 pathogenic trans-acting viral genes, meaning that cancer incidence is a
 direct consequence of the effects of proviral integration into host
 protooncogenes. The CA nucleic acid sequences can be used to diagnose
 leukaemia (especially breast cancer, prostate cancer, lymphoma or
 leukaemia) or a propensity to carcinoma by determination of the sequence
 of a CA gene, or by determination of CA gene expression in particular
 tissues. CA nucleic acids, proteins and antibodies are also useful as
 therapeutic agents and in screening and evaluating drug candidates. The
 present sequence represents a specifically claimed murine CA nucleic acid
 sequence of the invention. Note: The complete sequence data for this
 patent did not form part of the printed specification, but was obtained
 in electronic format directly from WIPO at

ftp.wipo.int/pub/published_pct_sequences.

Sequence 25032 BP, 6512 A, 5521 C, 5457 G, 6240 T, 0 U, 1302 Other;

Query Match 80.8%; Score 20.2; DB 9; Length 25032;
 Best Local Similarity 88.0%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGACAC 25
 Db 17554 AAAAAAAAAATCATTTATGTGACTC 17530

RESULT 10
 ADB72233/C
 ID ADB72233 standard; DNA; 25032 BP.

ADB72233;

04-DEC-2003 (first entry)

Mouse Myc gene.

mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
 cancer; neoplasm; adenocarcinoma; sarcoma; gene.

Mus sp.

MO2003008583-A2.

30-JAN-2003.

26-DEC-2001; 2001WO-US051291.

02-MAR-2001; 2001US-00798586.

23-OCT-2001; 2001US-00004113.

08-NOV-2001; 2001US-00052482.

30-NOV-2001; 2001US-00997722.

20-DEC-2001; 2001US-00034650.

(SAGR-) SAGRES DISCOVERY.

Morris DW, Engelhard EK;

WPI; 2003-239337/23.

Claim 1; SEQ ID NO 61; 2304pp; English.

The invention relates to a novel recombinant nucleic acid comprising a
 nucleotide sequence selected from any of the 660 sequences fully defined
 in the specification. A polynucleotide of the invention has cytostatic
 activity, and may have a use in gene therapy, or in a vaccine. The
 recombinant nucleic acids and polypeptides are useful for treating
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC sarcomas. The present sequence represents a mouse gene of the invention.

Sequence 25032 BP, 6512 A, 5521 C, 5457 G, 6240 T, 0 U, 1302 Other;

Query Match 80.8%; Score 20.2; DB 10; Length 25032;
 Best Local Similarity 88.0%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGCATGATTGTGACAC 25
 Db 17554 AAAAAAAAAATCATTTATGTGACTC 17530

RESULT 11
 ADB82935/C
 ID ADB82935 standard; DNA; 25032 BP.

XX AC ADE82935;
 XX XX
 DT 29-JAN-2004 (first entry)
 XX XX
 DE Mouse myc gene genomic DNA sequence.
 XX XX
 KW mouse; murine; cancer-associated nucleic acid; screening; cancer;
 XX lymphoma; leukemia; breast cancer; gene therapy; vaccine; ds.
 XX OS
 XX Mus sp.
 PN MO2003080808-A2.
 XX XX
 PD 02-OCT-2003.
 XX XX
 PF 21-MAR-2003; 2003MO-US008919.
 XX XX
 PR 21-MAR-2002; 2002US-0367025P.
 XX XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX XX
 PI Morris DW;
 XX XX
 DR MPI; 2003-865119/80.
 XX XX
 PT New cancer-associated proteins and nucleic acids, useful for screening
 PT for anticancer activity in a potential drug, or for detecting,
 PT diagnosing, preventing and treating cancers, e.g. lymphoma, leukemia or
 PT breast cancer.
 XX XX
 PS Disclosure; SEQ ID NO 19; 248bp; English.
 XX XX
 CC The invention comprises human and mouse cancer-associated nucleic acid
 CC sequences. The cancer associated nucleic acids of the invention are
 CC useful for screening for anticancer activity in a potential drug, as well
 CC as detecting, diagnosing, preventing and treating cancers (e.g. lymphoma,
 CC leukemia, or breast cancer). The present sequence represents a cancer-
 CC associated nucleic acid of the invention.
 XX XX
 SQ Sequence 25032 BP; 6512 A; 5521 C; 5457 G; 6240 T; 0 U; 1302 Other;
 Query Match 80.8%; Score 20.2; DB 10; Length 25032;
 Best Local Similarity 88.0%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
 Db 17554 AAAAAAAAAATCATTTGTGACTC 17530

PR 08-NOV-2001; 2001US-00052482.
 XX XX
 XX (SAGR-) SAGRES DISCOVERY.
 XX XX
 PI Morris DW, Engelhard EK;
 XX XX
 DR MPI; 2003-441462/41.
 XX XX
 PT New carcinoma associated nucleic acids and proteins, useful for screening
 PT drug candidates, or for diagnosing and treating carcinomas, e.g.
 PT lymphoma, breast cancer, prostate cancer or leukemia.
 XX XX
 XX Claim 1; SEQ ID NO 1; 793pp; English.
 XX XX
 PS This invention relates to novel recombinant nucleic acids for use in
 CC diagnosis and treatment of cancer, especially carcinomas, as well as the
 CC use of compositions in screening methods. The compositions of the
 CC invention may have cytostatic activity while the disclosed sequences may
 CC be useful for gene therapy. The carcinoma associated nucleic acids and
 CC proteins are useful for diagnosing and treating carcinomas, for example
 CC lymphoma, breast cancer, prostate cancer or leukemia, or for screening
 CC drug candidates or bioactive agents capable of binding to, or modulating
 CC the activity of, a carcinoma associated protein. The present sequence is
 CC the genomic DNA sequence of the mouse myc gene which is a carcinoma
 CC associated gene of the invention.
 XX XX
 SQ Sequence 25032 BP; 6513 A; 5521 C; 5456 G; 6240 T; 0 U; 1302 Other;
 Query Match 80.8%; Score 20.2; DB 10; Length 25032;
 Best Local Similarity 88.0%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
 Db 17554 AAAAAAAAAATCATTTGTGACTC 17530

RESULT 13
 AAD02697/c
 ID AAD02697 standard; DNA; 160552 BP.
 XX XX
 AC AAD02697;
 XX XX
 DT 02-MAY-2001 (first entry)
 XX XX
 DE Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
 XX XX
 KW Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy;
 KW selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ds.
 XX XX
 OS Homo sapiens.
 XX XX
 FH Key
 FH exon Location/Qualifiers
 FT 32847..32922
 FT /tag= a
 FT /number= 1
 FT /label= 4a.5U4
 FT 32923..35592
 FT /tag= b
 FT /cons_splice= (5'site:NO, 3'site:YBS)
 FT 35593..35674
 FT /tag= c
 FT /number= 2
 FT /label= 4a.5U3
 FT 35675..45093
 FT intron

XX 23-JUL-2001; 2001EP-00117608.
PR 24-MAY-2002; 2002EP-00011710.
XX
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
PI Heinrich G, Kerb R;
XX
DR WPI; 2003-289896/28.
XX
PT Use of irinotecan to treat cancer patient by determining if patient has
PT variant alleles of UGT1A1 gene, administering increased/decreased amounts
PT of irinotecan based on increased/decreased levels of UGT1A1 gene product.
XX
PS Disclosure; SEQ ID NO 683; 107pp; English.
XX
CC The invention relates to the novel use of irinotecan to treat a patient
CC suffering from cancer. This involves determining if the patient has one
CC or more variant alleles of the UGT1A1 gene, and if the patient has one or
CC more of such variant alleles, irinotecan is administered in an increased
CC or decreased amount in comparison to the amount that is administered
CC without regard to the patient's alleles in the UGT1A1 gene. The invention
CC has cytostatic activity. A composition of the invention acts as a
CC topoisomerase I inhibitor. The method is useful for treating a patient,
CC an animal e.g. mouse or a human, preferably African or Asian, suffering
CC from cancer such as colorectal, cervical, gastric cancer, lung, ovarian,
CC pancreatic cancer or malignant glioma. The present sequence is used in
CC the exemplification of the invention.
XX
SQ Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 10; Length 177380;
Best Local Similarity 91.3%; Pred. No. 4.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGCATGATTGTGAC 23
Db 96631 AAGAAAGAAAGCATGATTGTGAC 96653
RESULT 19
ADB96942
ID ADB96942 standard; DNA; 177380 BP.
XX
AC ADB96942;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human MDR1 related DNA sequence SEQ ID NO:683.
XX
KW irinotecan; colorectal cancer; cervical cancer; gastric cancer;
KW lung cancer; ovarian cancer; pancreatic cancer; malignant glioma;
KW multidrug resistance 1; MDR1; cytostatic; human; CYP3A5; MRP1; MDR1;
KW TOP1; ds.
XX
OS Homo sapiens.
XX
PI Heinrich G, Kerb R;
XX
PN WO2003013537-A2.
XX
PD 20-FEB-2003.
XX
PF 23-JUL-2002; 2002WO-EP008218.
XX
PR 23-JUL-2001; 2001EP-00117608.
PR 24-MAY-2002; 2002EP-00011710.
XX
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI Heinrich G, Kerb R;
XX
DR WPI; 2003-268145/26.
XX
PT New use of irinotecan for preparation of pharmaceutical compositions for

PT treating cancer in subject having genome with variant allele comprising
PT multidrug resistance 1 polynucleotide.
XX
PS Disclosure; SEQ ID NO 683; 130pp; English.
XX
CC The invention relates to the novel use of irinotecan or its derivative
CC for the preparation of pharmaceutical compositions for treating
CC colorectal, cervical, gastric, lung, ovarian or pancreatic cancer, or
CC malignant glioma in a subject having a genome with a variant allele which
CC comprises a multidrug resistance 1 (MDR1) polynucleotide. A composition
CC of the invention has cytostatic activity. The invention is useful for the
CC preparation of pharmaceutical compositions for treating colorectal,
CC cervical, gastric, lung, ovarian or pancreatic cancer, or malignant
CC glioma in a subject (preferably human, more preferably African or Asian)
CC or a mouse. The present sequence is used in the exemplification of the
CC invention.
XX
SQ Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 10; Length 177380;
Best Local Similarity 91.3%; Pred. No. 4.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGCATGATTGTGAC 23
Db 96631 AAGAAAGAAAGCATGATTGTGAC 96653
RESULT 20
ADB92133
ID ADB92133 standard; DNA; 177380 BP.
XX
AC ADB92133;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human MDR1 related DNA sequence SEQ ID NO:683.
XX
KW irinotecan; colorectal cancer; cervical cancer; gastric cancer;
KW lung cancer; ovarian cancer; pancreatic cancer; malignant glioma;
KW multidrug resistance 1; MDR1; cytostatic; human; UGT1A1; MRP1; TOP1; ds.
XX
OS Homo sapiens.
XX
PI Heinrich G, Kerb R;
XX
PN WO2003013535-A2.
XX
PD 20-FEB-2003.
XX
PF 23-JUL-2002; 2002WO-EP008220.
XX
PR 23-JUL-2001; 2001EP-00117608.
PR 24-MAY-2002; 2002EP-00011710.
XX
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX
PI Heinrich G, Kerb R;
XX
PN WPI; 2003-342400/32.
XX
PT New use of irinotecan for preparation of pharmaceutical compositions for
PT treating cancer in subject having genome with variant allele comprising
PT multidrug resistance 1 polynucleotide.
XX
PS Disclosure; SEQ ID NO 683; 104pp; English.
XX
CC The invention relates to a novel use of irinotecan or its derivative for
CC the preparation of a pharmaceutical composition for treating colorectal,
CC cervical, gastric, lung, ovarian or pancreatic cancer, or malignant
CC glioma in a subject having a genome with a variant allele which comprises
CC a multidrug resistance 1 (MDR1) polynucleotide. A composition of the
CC invention has cytostatic activity. The present sequence is used in the
CC exemplification of the invention.
XX

SEQ Sequence 177380 BP; 58338 A; 32530 C; 31389 G; 55123 T; 0 U; 0 Other;
Query Match 79.2%; Score 19.8; DB 10; Length 177380;
Best Local Similarity 91.3%; Pred. No. 4.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTGAC 23
DB 96631 AAGAAAGAAAGCATGATTGTGAC 96653
RESULT 21
ABL81814
ID ABL81814 standard; cDNA, 329 BP.
XX ABL81814;
XX
XX
XX 17-MAY-2002 (first entry)
XX
XX Human ovarian cancer related cDNA clone SEQ ID NO:4792.
XX
XX Human; ovarian cancer; ovarian tumour; cytosstatic; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200192581-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US017756.
XX
XX 26-MAY-2000; 2000US-0207484P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Algate PA, Harlocker SL, Jones R;
XX
XX WPI; 2002-122075/16.
XX
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
XX Claim 1; SEQ ID NO 4792; 489pp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) having a cDNA sequence
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (SI)
CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridizes to (SI) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridizing to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridizing to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
XX Sequence 329 BP; 121 A; 57 C; 58 G; 93 T; 0 U; 0 Other;
SQ
Query Match 77.6%; Score 19.4; DB 6; Length 329;
Best Local Similarity 95.2%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTG 21
DB 68 AAAAAAAAAAGCATGATTGTG 88
RESULT 22
ABL64629
ID ABL64629 standard; DNA, 356 BP.
XX ABL64629;
XX
XX 15-MAY-2002 (first entry)
XX
XX Stomach cancer related gene sequence SEQ ID NO:2966.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; wilm's tumour; adenocarcinoma;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US010838.
XX
XX 05-JUN-2000; 2000US-0209473P.
XX 05-JUN-2000; 2000US-0209531P.
XX 18-SEP-2000; 2000US-0233133P.
XX 18-SEP-2000; 2000US-0233617P.
XX 20-SEP-2000; 2000US-0234009P.
XX 20-SEP-2000; 2000US-0234034P.
XX 20-SEP-2000; 2000US-0234052P.
XX 22-SEP-2000; 2000US-0234509P.
XX 22-SEP-2000; 2000US-0234567P.
XX 25-SEP-2000; 2000US-0234923P.
XX 25-SEP-2000; 2000US-0234924P.
XX 25-SEP-2000; 2000US-0235077P.
XX 25-SEP-2000; 2000US-0235082P.
XX 25-SEP-2000; 2000US-0235134P.
XX 25-SEP-2000; 2000US-0235280P.
XX 26-SEP-2000; 2000US-0235637P.
XX 26-SEP-2000; 2000US-0235638P.
XX 27-SEP-2000; 2000US-0235711P.
XX 27-SEP-2000; 2000US-0235720P.
XX 27-SEP-2000; 2000US-0235840P.
XX 27-SEP-2000; 2000US-0235863P.
XX 28-SEP-2000; 2000US-0236028P.
XX 28-SEP-2000; 2000US-0236032P.
XX 28-SEP-2000; 2000US-0236033P.
XX 28-SEP-2000; 2000US-0236034P.
XX 28-SEP-2000; 2000US-0236109P.
XX 28-SEP-2000; 2000US-0236111P.
XX 29-SEP-2000; 2000US-0236842P.
XX 29-SEP-2000; 2000US-0236891P.
XX 02-OCT-2000; 2000US-0237172P.
XX 02-OCT-2000; 2000US-0237173P.
XX 02-OCT-2000; 2000US-0237278P.
XX 02-OCT-2000; 2000US-0237294P.
XX 02-OCT-2000; 2000US-0237295P.
XX 02-OCT-2000; 2000US-0237316P.
XX 02-OCT-2000; 2000US-0237425P.
XX 03-OCT-2000; 2000US-0237598P.
XX 03-OCT-2000; 2000US-0237604P.
XX 03-OCT-2000; 2000US-0237606P.
XX 03-OCT-2000; 2000US-0237608P.
XX 01-NOV-2000; 2000US-0244867P.
XX 01-NOV-2000; 2000US-0245084P.
XX
XX (AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.
 DR
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 XX
 PS Claim 1; SEQ ID NO 2966; 44pp; English.
 CC
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour
 CC
 XX
 SQ Sequence 356 BP; 133 A; 63 C; 62 G; 98 T; 0 U; 0 Other;
 Query Match 77.6%; Score 19.4; DB 6; Length 356;
 Best Local Similarity 95.2%; Pred. No. 3.8e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTG 21
 Db 68 AAAAAAAAAAGCAAGATTGTG 88
 RESULT 23
 ABL80674
 ID ABL80674 standard; cDNA; 356 BP.
 AC ABL80674;
 XX
 DT 17-MAY-2002 (first entry)
 XX
 DE Human ovarian cancer related cDNA clone SEQ ID NO:3652.
 XX
 KM Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200192581-A2.
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US017756.
 PR 26-MAY-2000; 2000US-0207484P.
 XX
 PA (CORI-) CORIXA CORP.
 PI Algate PA, Harlocker SL, Jones R;
 XX WPI; 2002-122075/16.
 DR
 XX
 PT Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.
 XX

PS Claim 1; SEQ ID NO 3652; 489pp; English.
 XX
 CC The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
 CC or antigen presenting cells that express (II). (I) has cytostatic
 CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 CC detecting ovarian cancer in a patient's biological sample preferably
 CC serum or ovarian tissue. The method comprises contacting a biological
 CC sample from a patient with (IV), detecting the amount of polynucleotide
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff
 CC value and thereby detecting ovarian cancer in the patient, where the
 CC amount of polynucleotide hybridising to (IV) is detected preferably by
 CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
 CC useful for stimulating and/or expanding T cells specific for an ovarian
 CC tumour protein comprising contacting T cells with (III) or (II). (III) is
 CC useful in design and preparation of ribozyme molecules for inhibiting
 CC expression of the tumour polypeptides and proteins in tumour cells; and
 CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
 CC library using well known techniques
 CC
 XX
 SQ Sequence 356 BP; 133 A; 63 C; 62 G; 98 T; 0 U; 0 Other;
 Query Match 77.6%; Score 19.4; DB 6; Length 356;
 Best Local Similarity 95.2%; Pred. No. 3.8e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTG 21
 Db 68 AAAAAAAAAAGCAAGATTGTG 88
 RESULT 24
 ABL87666
 ID ABL87666 standard; cDNA; 377 BP.
 AC ABL87666;
 XX
 DT 17-MAY-2002 (first entry)
 XX
 DE Human ovarian cancer related cDNA clone SEQ ID NO:10644.
 XX
 KM Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200192581-A2.
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US017756.
 PR 26-MAY-2000; 2000US-0207484P.
 XX
 PA (CORI-) CORIXA CORP.
 PI Algate PA, Harlocker SL, Jones R;
 XX WPI; 2002-122075/16.
 DR
 XX
 PT Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.
 XX
 PS Claim 1; SEQ ID NO 10644; 489pp; English.
 CC
 CC The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 CC (III) having a sequence (S2), a T cell population of (II),
 CC or antigen presenting cells that express (II). (I) has cytostatic
 CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 CC detecting ovarian cancer in a patient's biological sample preferably
 CC serum or ovarian tissue. The method comprises contacting a biological
 CC sample from a patient with (IV), detecting the amount of polynucleotide
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff
 CC value and thereby detecting ovarian cancer in the patient, where the
 CC amount of polynucleotide hybridising to (IV) is detected preferably by
 CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
 CC useful for stimulating and/or expanding T cells specific for an ovarian
 CC tumour protein comprising contacting T cells with (III) or (II). (III) is
 CC useful in design and preparation of ribozyme molecules for inhibiting
 CC expression of the tumour polypeptides and proteins in tumour cells; and
 CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
 CC library using well known techniques

CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques

XX
SQ Sequence 377 BP; 137 A; 67 C; 66 G; 106 T; 0 U; 1 Other;

Query Match 77.6%; Score 19.4; DB 6; Length 377;
Best Local Similarity 95.2%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTG 21
DB 68 AAAAAAAAAAGCAAGATTGTG 88

RESULT 25
ABL82013/c
ID ABL82013 standard; cDNA; 393 BP.
XX
AC ABL82013;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:4991.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US017756.
XX
PR 26-MAY-2000; 2000US-0207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
PT MPI; 2002-122075/16.
XX
DR
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
PS Claim 1; SEQ ID NO 4991; 4899P; English.
XX
XX The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide

CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques

XX
SQ Sequence 393 BP; 108 A; 78 C; 67 G; 140 T; 0 U; 0 Other;

Query Match 77.6%; Score 19.4; DB 6; Length 393;
Best Local Similarity 95.2%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTG 21
DB 342 AAAAAAAAAAGCAAGATTGTG 322

RESULT 26
ABL87751/c
ID ABL87751 standard; cDNA; 396 BP.
XX
AC ABL87751;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:10729.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US017756.
XX
PR 26-MAY-2000; 2000US-0207484P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Algate PA, Harlocker SL, Jones R;
XX
PT MPI; 2002-122075/16.
XX
DR
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
PS Claim 1; SEQ ID NO 10729; 4899P; English.
XX
XX The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is

CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 396 BP; 107 A; 79 C; 67 G; 143 T; 0 U; 0 Other;
Query Match 77.6%; Score 19.4; DB 6; Length 396;
Best Local Similarity 95.2%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTG 21
DB 343 AAAAAAAAAAGCATGATTGTG 323
RESULT 27
AAH1468/c
ID AAH1468 standard; cDNA; 553 BP.
XX
AC AAH1468;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:8303.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 3; SEQ ID NO 8303; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 553 BP; 177 A; 96 C; 119 G; 157 T; 0 U; 4 Other;
Query Match 77.6%; Score 19.4; DB 4; Length 553;
Best Local Similarity 95.2%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTG 21
DB 28 AAAAAAAAAAGCATGATTGTG 8
RESULT 28
AAH14440
ID AAH14440 standard; cDNA; 3434 BP.
XX
AC AAH14440;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11907.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 11907; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH1629 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX SQ Sequence 3434 BP; 811 A; 1061 C; 773 G; 789 T; 0 U; 0 Other;

Query Match 77.6%; Score 19.4; DB 4; Length 3434;
Best Local Similarity 95.2%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 3407 AAAAAAAAAAGCCTGATTGTG 3427

RESULT 29

ADV35067
ID ADV35067 standard; cDNA; 3434 BP.

XX AC ADV35067;

XX DT 10-FEB-2005 (first entry)

XX DE Human cDNA differentially expressed in the presence of valproate Seq143.

XX KW human; valproate; ss; multi-parameter high throughput screening; MPHTS;

XX KW disease signature; neuropsychiatric; neurodegenerative; schizophrenia;

XX KW bipolar affective disorder; BAD; autism; Parkinson's;

XX KW Alzheimer's disease; neuroleptic; nootropic; antianemic; antidepressant.

XX OS Homo sapiens.

XX PN US2003096264-A1.

XX PD 22-MAY-2003.

XX PF 18-JUN-2002; 2002US-00175523.

XX PR 18-JUN-2001; 2001US-0299151P.

XX PR 07-SEP-2001; 2001US-0317828P.

XX PR 25-SEP-2001; 2001US-0325150P.

XX PR 14-NOV-2001; 2001US-033047P.

XX PR 18-JAN-2002; 2002US-0349936P.

XX PR 04-MAR-2002; 2002US-0361834P.

XX PA (PSTC-) PSYCHIATRIC GENOMICS INC.

XX PI Altar CA, Brockman JA, Evans D, Hook D, Klimczak LJ, Laeng P,

XX PI Palfreyman M, Rajan P;

XX DR WPI; 2004-118903/12.

XX PS Claim 12; SEQ ID NO 143; 39pp; English.

XX This invention relates to a novel screening method identified as a multi-
CC parameter high throughput screening (MPHTS) assay. Specifically, it
CC refers to an assay that utilizes the disease signature of a plurality of
CC different genes associated with a particular disease, and identifies
CC differential expression between those cells taken from individuals
CC affected by that disease and those that are not affected. The present
CC invention then describes the screening of candidate pharmaceutical
CC compounds to identify those that have a potential therapeutic benefit for
CC the treatment of neuropsychiatric and neurodegenerative disorders
CC including schizophrenia, bipolar affective disorder (BAD) and autism, as
CC well as Parkinson's and Alzheimer's disease. Accordingly, the compounds

CC of this invention exhibit various activities including neuroleptic,
CC nootropic, antianemic and antidepressant. Furthermore, the screening
CC method used in MPHTS will be automated, such that a large number of test
CC compounds may be rapidly screened with a minimal amount of labour and
CC effort. This polynucleotide is a human cDNA sequence of a gene that is
CC differentially expressed in the presence of the therapeutic compound
CC valproate, given in an exemplification of the invention.

XX SQ Sequence 3434 BP; 811 A; 1061 C; 773 G; 789 T; 0 U; 0 Other;

Query Match 77.6%; Score 19.4; DB 13; Length 3434;
Best Local Similarity 95.2%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTG 21
Db 3407 AAAAAAAAAAGCCTGATTGTG 3427

RESULT 30

ABD33098/c
ID ABD33098 standard; DNA; 219352 BP.

XX AC ABD33098;

XX DT 18-NOV-2004 (first entry)

XX DE Murine cancer-associated (CA) gene MD07-008.

XX KW Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;

XX KW ds; cancer; cytostatic.

XX OS Mus musculus.

XX PN WO2004058146-A2.

XX PD 15-JUL-2004.

XX PF 15-DEC-2003; 2003WO-US040081.

XX PR 17-DEC-2002; 2002US-00322281.

XX PA (SAGR-) SAGRES DISCOVERY INC.

XX PI Morris DW, Malandro MS;

XX DR WPI; 2004-499109/47.

XX PT Novel human cancer associated protein encoded within open reading frame

XX PS Of cancer associated gene, useful as targets for diagnosing cancer.

XX PS Disclosure; SEQ ID NO 45; 182pp; English.

XX This invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a murine CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 219352 BP; 57476 A; 45974 C; 48501 G; 66395 T; 0 U; 1006 Other;

Query Match 77.6%; Score 19.4; DB 13; Length 219352;

Best Local Similarity 95.2%; Pred. No. 6.1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACATGATTGTG 21
|||||

Db 165651 AAAAAAAAAATGATGATTGTG 165631

RESULT 31

ACN48867/c
ID ACN48867 standard; cDNA; 163 BP.

AC ACN48867;

XX 02-DEC-2004 (first entry)

DE Cotton primed seed EST Clone ID: LIB3825-028-06-K6-A2, SEQ:3648.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;

KM variety DP50B; library LIB3825; molecular tag; molecular marker;

KM genetic mapping; molecular mapping; seed germination; plant growth;

XX plant quality; plant yield; plant breeding; tissue printing; ss.

OS Gosypium hirsutum.

XX US2004123340-A1.

PD 24-JUN-2004.

XX 12-DEC-2001; 2001US-00021323.

XX 14-DEC-2000; 2000US-0255619P.

PA (DEIK/) DEIKMAN J.
PA (FENG/) FENG P C C.
PA (FINC/) FINCHER K L.
PA (ZIEG/) ZIEGLER T E.

PI Deikman J, Feng PCC, Fincher KL, Ziegler TE;

DR WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its
PT fragment, useful for isolating a variety of agronomically significant
PT genes associated with plant growth, quality or yield, and as molecular
PT tags to map genes.

XX Claim 1; SEQ ID NO 3648; 34pp; English.

CC The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC from primed or non-primed seeds from variety DP50B, mature seeds from
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC tissue, developing fibres, carpel walls and septa from variety
CC Nucleon33B. The invention also relates to substantially purified
CC proteins or their fragments encoded by nucleic acid molecules of the
CC invention, and to transformed plants having a nucleic acid construct
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC molecular tags to isolate genetic regions, to isolate genes, to map
CC genes, to determine gene function and to determining whether genes are
CC members of a particular gene family. The nucleic acid molecules may be
CC used for isolating a variety of agronomically significant genes
CC associated with plant growth, quality, yield, and could also serve as
CC links in metabolic and catabolic pathways. The nucleic acid molecules are
CC also useful for identifying genes important in initiating and maintaining
CC seed germination or that may be used to mitigate stresses encountered
CC during seed germination. The ESTs additionally enable the acquisition of
CC promoters and cis-regulatory elements which will be useful to express
CC agronomically significant genes in these tissues and/or other tissues,
CC and also permits the acquisition of molecular markers useful in breeding
CC schemes, genetic and molecular mapping, and in cloning of agronomically
CC significant genes. The nucleic acid molecules are further useful for
CC detecting the expression level or pattern of a protein or mRNA and for

CC detecting the presence or quantity of a protein by tissue printing. The
CC present sequence represents a specifically claimed EST isolated from a
CC cotton variety DP50B primed seed cDNA library (LIB3825). The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from the US Patent office at
CC seqdata.uspto.gov/sequence.html?docid=US20040123340
CC
XX
SQ Sequence 163 BP; 32 A; 24 C; 35 G; 72 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 13; Length 163;
Best Local Similarity 87.5%; Pred. No. 4.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACATGATTGACA 24
|||||

Db 130 AAAAAAAAAACATGATTAGACA 107

RESULT 32

ACN48946
ID ACN48946 standard; cDNA; 166 BP.

AC ACN48946;

XX 02-DEC-2004 (first entry)

DE Cotton primed seed EST Clone ID: LIB3825-028-06-N6-A2, SEQ:3727.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;

KM variety DP50B; library LIB3825; molecular tag; molecular marker;

KM genetic mapping; molecular mapping; seed germination; plant growth;

XX plant quality; plant yield; plant breeding; tissue printing; ss.

OS Gosypium hirsutum.

XX US2004123340-A1.

PD 24-JUN-2004.

XX 12-DEC-2001; 2001US-00021323.

XX 14-DEC-2000; 2000US-0255619P.

PA (DEIK/) DEIKMAN J.
PA (FENG/) FENG P C C.
PA (FINC/) FINCHER K L.
PA (ZIEG/) ZIEGLER T E.

PI Deikman J, Feng PCC, Fincher KL, Ziegler TE;

DR WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its
PT fragment, useful for isolating a variety of agronomically significant
PT genes associated with plant growth, quality or yield, and as molecular
PT tags to map genes.

XX Claim 1; SEQ ID NO 3727; 34pp; English.

CC The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC from primed or non-primed seeds from variety DP50B, mature seeds from
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC tissue, developing fibres, carpel walls and septa from variety
CC Nucleon33B. The invention also relates to substantially purified
CC proteins or their fragments encoded by nucleic acid molecules of the
CC invention, and to transformed plants having a nucleic acid construct
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC molecular tags to isolate genetic regions, to isolate genes, to map
CC genes, to determine gene function and to determining whether genes are
CC members of a particular gene family. The nucleic acid molecules may be
CC used for isolating a variety of agronomically significant genes
CC associated with plant growth, quality, yield, and could also serve as

CC links in metabolic and catabolic pathways. The nucleic acid molecules are
CC also useful for identifying genes important in initiating and maintaining
CC seed germination or that may be used to mitigate stresses encountered
CC during seed germination. The ESTs additionally enable the acquisition of
CC promoters and cis-regulatory elements which will be useful to express
CC agronomically significant genes in these tissues and/or other tissues,
CC and also permits the acquisition of molecular markers useful in breeding
CC schemes, genetic and molecular mapping, and in cloning of agronomically
CC significant genes. The nucleic acid molecules are further useful for
CC detecting the expression level or pattern of a protein or mRNA and for
CC detecting the presence or quantity of a protein by tissue printing. The
CC present sequence represents a specifically claimed EST isolated from a
CC cotton variety Dpsb primed seed cDNA library (LIB3825). The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?docID=US20040123340
XX
SQ Sequence 166 BP; 71 A; 37 C; 24 G; 34 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 13; Length 166;
Best Local Similarity 87.5%; Pred. No. 4,4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATGTGACA 24
Db 32 AAAAAAAAAACGATGATTAGACA 55

RESULT 33
ID AAL00715 standard; cDNA; 414 BP.
XX AAL00715;
AC
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen cDNA SEQ ID NO: 716.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US001339.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216860P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226779P.
PR 22-AUG-2000; 2000US-0226811P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227018P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228987P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234232P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240560P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.

```
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 03-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX DR P-PSDB; AAM94745.
XX
XX PT Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX PS Claim 1; SEQ ID NO 716; 1297pp + Sequence Listing; English.
XX
XX CC The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a coding sequence of the
XX invention
XX
XX SQ Sequence 414 BP; 140 A; 94 C; 69 G; 107 T; 0 U; 4 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 4; Length 414;
XX Best Local Similarity 87.5%; Pred. No. 4.7e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAAGCATGATTGTGACA 24
XX ||||| ||||| ||||| |||||
XX Db 135 AAAAAAAAAATCAGATTGTGAAA 158
XX
XX RESULT 34
XX AA114617
XX ID AA114617 standard; DNA; 464 BP.
XX
```

```
AC AA114617;
XX
XX DT 12-OCT-2001 (first entry)
XX
XX DE Probe #4550 for gene expression analysis in human cervical cell sample.
XX
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157278-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000670.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-468901/53.
XX
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX
XX PT Claim 25; SEQ ID NO 4550; 487bp; English.
XX
XX PS The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging of
XX diseases of the cervix, notably cervical cancer. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 4; Length 464;
XX Best Local Similarity 87.5%; Pred. No. 4.7e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAAGCATGATTGTGACA 24
XX ||||| ||||| ||||| |||||
XX Db 137 AAAAAAAAAACATGATTGTGAAA 160
XX
XX RESULT 35
XX ABA56346
XX ID ABA56346 standard; DNA; 464 BP.
XX
XX AC ABA56346;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human foetal liver single exon nucleic acid probe #4651.
XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157277-A2.
XX
```

XX PD 09-AUG-2001.
 XX XD
 XX PI 30-JAN-2001; 2001WO-US000669.
 XX PF
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human fetal liver.
 XX
 PS Claim 1; SEQ ID NO 4651; 639pp + Sequence Listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human fetal liver. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
 XX
 QY Query Match 76.8%; Score 19.2; DB 4; Length 464;
 Best Local Similarity 87.5%; Pred. No. 4.7e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 XX
 Db 1 AAAAAAAAAAGCATGATTGTGACA 24
 137 AAAAAAAAAACCATGATTGTGAAA 160
 XX
 RESULT 36
 AAI35989 standard; DNA; 464 BP.
 XX
 AC AAI35989;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #4675 used to measure gene expression in human placenta sample.
 XX
 KM Probe; microarray; human; placenta; antenatal diagnosis;
 KM genetic disorder; ss.
 XX
 OS Homo sapiens.
 OS
 PN MO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488997/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human placenta.
 XX
 PS Claim 25; SEQ ID NO 4675; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders
 XX
 SQ Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
 XX
 QY Query Match 76.8%; Score 19.2; DB 4; Length 464;
 Best Local Similarity 87.5%; Pred. No. 4.7e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 XX
 Db 1 AAAAAAAAAAGCATGATTGTGACA 24
 137 AAAAAAAAAACCATGATTGTGAAA 160
 XX
 RESULT 37
 ABA45831 standard; DNA; 464 BP.
 XX
 AC ABA45831;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #4526.
 XX
 KM Human; microarray; single exon probe; gene expression; breast; disease;
 KM cancer; ss.
 XX
 OS Homo sapiens.
 OS
 PN MO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000662.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes, useful
 PT for measuring gene expression in sample derived from human breast,
 PT comprises number of single exon nucleic acid probes.
 XX
 PS Claim 1; SEQ ID NO 4526; 327pp + Sequence Listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting the
 CC probes with a collection of detectably labelled nucleic acids derived

CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC
XX Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
SQ
Query Match 76.8%; Score 19.2; DB 4; Length 464;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACATGATTGTGACA 24
Db 137 AAAAAAAAAACATGATTGTGAAA 160
RESULT 38
ABA25987
ID ABA25987 standard; DNA; 464 BP.
XX
AC ABA25987;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #4453 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KM cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
XX WO200157274-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000666.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488899/53.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
PT
XX
XX Claim 1; SEQ ID NO 4453; 530pp; English.
PS
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,

CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
CC
XX Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
SQ
Query Match 76.8%; Score 19.2; DB 4; Length 464;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACATGATTGTGACA 24
Db 137 AAAAAAAAAACATGATTGTGAAA 160
RESULT 39
AAK30026
ID AAK30026 standard; DNA; 464 BP.
XX
AC AAK30026;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 4583.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KM microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000668.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488900/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
PT
XX
XX Example 4; SEQ ID NO 4583; 658pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention
CC
XX
XX Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
SQ
Query Match 76.8%; Score 19.2; DB 4; Length 464;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACATGATTGTGACA 24
Db 137 AAAAAAAAAACATGATTGTGAAA 160

```

RESULT 40
AAK04525
ID AAK04525 standard; DNA; 464 BP.
XX
AC AAK04525;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 4516.
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
XX ss.
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX MPI; 2001-483446/52.
XX
DR Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 4516; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is one of the probes of the
XX CC invention
XX
SQ Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
XX
Query Match 76.8%; Score 19.2; DB 4; Length 464;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 137 AAAAAAAAAACATGATTGTGAAA 160
XX
RESULT 41
ABS29678
ID ABS29678 standard; DNA; 464 BP.
XX
AC ABS29678;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID NO 4668.
XX
XX Human; single exon nucleic acid probe; liver; cirrhosis;

```

```

XX hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
XX coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX MPI; 2001-488698/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
XX Claim 1; SEQ ID NO 4668; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX CC (I) may be used for predicting, measuring and displaying gene expression
XX CC in samples derived from human adult liver. The genes identified may be
XX CC involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX CC associated with coronary heart disease. ABS25011-ABS51005 represent human
XX CC liver single exon nucleic acid probes of the invention. Note: The
XX CC sequence information for this patent does not appear in the printed
XX CC specification but was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
XX
Query Match 76.8%; Score 19.2; DB 4; Length 464;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATGATTGTGACA 24
Db 137 AAAAAAAAAACATGATTGTGAAA 160
XX
RESULT 42
ABS04598
ID ABS04598 standard; DNA; 464 BP.
XX
AC ABS04598;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID NO 4589.
XX
XX Human; de; single exon probe; asthma; lung cancer; COPD; ILD;
XX XX chronic obstructive pulmonary disease; interstitial lung disease;
XX XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX XX Heremans-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
XX XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

```

KM primary ciliary dyskinesia; pulmonary hypertension;
 KM hyaline membrane disease.
 OS Homo sapiens.
 XX MO200186003-A2.
 PN
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000665.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 DR
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples.
 XX
 PS Claim 1; SEQ ID NO 4589; 634pp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA; and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 CC Karsagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a single exon probe of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 464 BP; 166 A; 81 C; 67 G; 150 T; 0 U; 0 Other;
 Query Match 76.8%; Score 19.2; DB 6; Length 464;
 Best local Similarity 87.5%; Pred. No. 4.7e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 AAAAAAAAAAGCATGATGGACA 24
 Db 137 AAAAAAAAAACATGATTTGAAA 160
 RESULT 43
 ACH77614/C
 ID ACH77614 standard; DNA; 516 BP.
 XX
 AC ACH77614;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Human genome derived single exon probe #10809.
 XX
 KM Human, probe; ss; gene expression; single exon probe; microarray;
 KM alternative splicing event; genomic alteration.
 XX
 OS Homo sapiens.
 XX
 PN US2003194704-A1.
 PD 16-OCT-2003.
 XX
 PF 03-APR-2002; 2002US-00029386.
 XX
 PR 03-APR-2002; 2002US-00029386.
 XX
 PA (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX
 PI Penn SG, Rank DR, Hanzel DK;
 XX WPI; 2004-119264/12.
 DR
 XX
 PT New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX
 PS Claim 15; SEQ ID NO 10809; 80pp; English.
 XX
 CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridises under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC a method of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subsequence, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterising
 CC alternative splicing events, in detecting and characterising gross

CC products and methods of the invention can be used for the diagnosis,
CC prognosis, and treatment of cancer, tumor progression,
CC hyperproliferative cell growth, and accompanying physical and biological
CC manifestations. They can be used particularly for prostatic disorders
CC such as metastatic prostate cancer, localized prostate cancer, or benign
CC prostate hyperplasia (BPH)
XX
SQ Sequence 782 BP; 181 A; 201 C; 171 G; 223 T; 0 U; 6 Other;
Query Match 76.8%; Score 19.2; DB 3; Length 782;
Best Local Similarity 87.5%; Pred. No. 4.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
DY 1 AAAAAAAAAAGCATGTTGACA 24
Db 469 AAAAAAAAAAGATTGTGACA 492
RESULT 46
ADT17668
ID ADT17668 standard; cDNA; 1654 BP.
AC ADT17668;
XX
DT 13-JAN-2005 (first entry)
XX
DE Plant cDNA, Seq ID 2994.
XX
KW Plant; ss; gene; transgenic; cold tolerance; growth rate;
KW drought tolerance; disease resistance; galactomannan production;
KW plant growth regulator; heat tolerance; herbicide tolerance;
KW lignin production; extreme osmotic condition tolerance;
KW pathogens resistance; pest resistance; yield improvement; seed oil yield;
KW seed protein yield.
XX
XX Viridiplantae.
OS
XX
PN US004216190-A1.
XX
PD 28-OCT-2004.
XX
PE 18-DEC-2003; 2003US-00739930.
XX
PR 28-APR-2003; 2003US-00424599.
XX
PR 28-APR-2003; 2003US-00425115.
XX
XX (KOVA/) KOVALIC D K.
PA
XX
PI KOvalic DK;
XX
DR WPI; 2004-757369/74.
XX
XX
PT New recombinant DNA constructs useful in the field of biochemistry and
PT genetics, and in particular for producing transgenic plants with improved
PT biological characteristics.
XX
XX Claim 1; SEQ ID NO 2994; 14pp; English.
XX
XX The invention relates a recombinant DNA construct comprising a
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,
CC Arabidopsis, wheat and rape but the specification does not indicate which
CC sequences is derived from which organism. Also included is a method of
CC producing a plant having an improved property, comprising transforming a
CC plant with a recombinant DNA construct comprising a promoter region
CC functional in a plant cell operably joined to a polynucleotide encoding a
CC polypeptide associated with the property, and growing the transformed
CC plant. The property is selected from improving plant cold tolerance, for
CC manipulating growth rate in plant cells by modification of the cell cycle
CC pathway, for improving plant drought tolerance, for providing increased
CC resistance to plant disease, for galactomannan production, for production
CC of plant growth regulators, for improving plant heat tolerance, for

CC improving plant tolerance to herbicides, for increasing the rate of
CC homologous recombination in plants, for lign production, for improving
CC plant tolerance to extreme osmotic conditions, for improving plant
CC tolerance to pathogens or pests, for yield improvement by modification of
CC photosynthesis, for modifying seed oil yield and/or content, for
CC modifying seed protein yield and/or content, for yield improvement by
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake
CC and for yield improvement by providing improved plant growth and
CC development under at least one stress condition. The polynucleotide may
CC also encode a plant transcription factor. The methods and compositions of
CC the present invention are useful in the field of biochemistry and
CC genetics, in particular for producing transgenic plants with improved
CC biological characteristics such as increased yield, improved nitrogen
CC flow, increasing plant tolerance to cold or heat, improving plant
CC tolerance to extreme osmotic and drought conditions, and improving plant
CC tolerance to plant pests or pathogens. They can also be used in physical
CC arrays of molecules, plant breeding markers, computer-based storage and
CC analysis systems. The present sequence is one of the 5544 plant cDNA
CC sequences of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPRO at
CC seqdata.uspro.gov/sequence.html?docid=20040216190.
XX
SQ Sequence 1654 BP; 387 A; 436 C; 442 G; 389 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 13; Length 1654;
Best Local Similarity 87.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
DY 1 AAAAAAAAAAGCATGTTGACA 24
Db 1560 AAAAAAAAAAAATTTATTTGACA 1583
RESULT 47
AAH16416/C
ID AAH16416 standard; cDNA; 1928 BP.
XX
AC AAH16416;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15395.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
XX
PR 27-AUG-1999; 99JP-00300253.
XX
PR 11-JAN-2000; 2000JP-00118776.
XX
PR 02-MAY-2000; 2000JP-00183767.
XX
PR 09-JUN-2000; 2000JP-00241899.
XX
XX
PA (HELI-) HELIX RES INST.
XX
XX
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 15395; 2537pp + Sequence Listing; English.
PS

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 1928 BP; 558 A; 285 C; 317 G; 768 T; 0 U; 0 Other;

SO Query Match 76.8%; Score 19.2; DB 4; Length 1928; Best Local Similarity 87.5%; Pred. No. 5.2e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTGTGAGACA 24
Db 564 AAAAAAAAAACATGTGTGAGACA 541

RESULT 48
ADA71787
ID ADA71787 standard; DNA; 2000 BP.

XX ADA71787;

XX 20-NOV-2003 (first entry)

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 5112.

XX KM Plant; bacterial infection; fungal infection; viral infection; rice; gene; ds.

XX OS Oryza sativa.

XX PN WO2003000898-A1.

XX PD 03-JAN-2003.

XX PF 22-JUN-2001, 2001WO-IB001105.

XX PR 22-JUN-2001, 2001WO-IB001105.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y, Katsirli F, Qian S, Tao Y, Whitnam S, Xie Z, Zhu T, Zou G, WPI; 2003-175290/17.

XX DR WPI; 2003-175290/17.

XX PT Identifying at least one gene involved in plant resistance or response to pathogenic infection for conferring resistance or tolerance to a plant to bacterial, fungal or viral infection by determining or detecting plant gene expression.

XX PS Claim 27; SEQ ID NO 5112; 899bp; English.

XX The present invention relates to a method (M1) for identifying genes

CC involved in plant resistance or response to pathogenic infection. M1 comprises identifying a gene whose expression is significantly altered in the incompatible interaction of plant gene expression relative to expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interaction. (M1) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to illustrate the invention.

XX Sequence 2000 BP; 676 A; 388 C; 347 G; 589 T; 0 U; 0 Other;

SO Query Match 76.8%; Score 19.2; DB 8; Length 2000; Best Local Similarity 87.5%; Pred. No. 5.2e+02; Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGTGTGAGACC 25
Db 366 AAAAAAAAAAGCATGTGTGAGACC 389

RESULT 49
AAF77275/c
ID AAF77275 standard; DNA; 2366 BP.

XX AAF77275;

XX 22-MAY-2001 (first entry)

XX DE Human alpha-myosin heavy chain (MHC) promoter DNA sequence.

XX KM Human; cardiac helix-loop-helix factor; CHF; transcription factor; cardiomyocyte; proliferation; myocarditis; myocardial infarction; KM cardiomyocyte regeneration; angiogenesis inhibitor; differentiation; smooth muscle cell growth; heart disease; MHC; promoter; ds.

XX OS Homo sapiens.

XX PN WO200112126-A2.

XX PD 22-FEB-2001.

XX PF 11-AUG-2000; 2000WO-US021858.

XX PR 13-AUG-1999; 99US-0148974P.

XX PA (HARD) HARVARD COLLEGE.

XX PI Lee M, Chin MT; WPI; 2001-211119/21.

XX DR WPI; 2001-211119/21.

XX PT Novel cardiovascular helix-loop-helix factor polypeptide and polynucleotide useful for regenerating heart tissue and promoting smooth muscle differentiation to treat heart disease or heart injury.

XX PS Disclosure; Page 22-23; 46pp; English.

XX This invention relates to human cardiovascular helix-loop-helix factor (CHF) polypeptides. CHF is a transcription factor expressed in cardiovascular tissue, which contains a basic helix-loop-helix domain. CHF-1 shares structural homology with hairy, a Drosophila protein essential for the development of the peripheral nervous system. The CHF polypeptide is useful for expanding an endogenous population of proliferative cardiomyocytes in a heart tissue, promoting proliferation of a cardiomyocyte in a mammal having myocarditis or that which has suffered myocardial infarction and for regenerating cardiomyocytes in vivo. DNA encoding a CHF-1 polypeptide is also useful for promoting proliferation of a cardiomyocyte, regenerating cardiomyocytes in vitro, inhibiting angiogenesis in a tissue, inducing differentiation of smooth muscle cells in a mammalian tissue, preferably venous tissue, inducing growth of smooth muscle cells in a vein explant, promoting smooth muscle

cell regeneration in an injured or diseased vascular tissue (venous or arterial) and for reducing vein graft stenosis in a mammal by contacting the tissue ex vivo with the DNA prior to implantation of the tissue into an artery of the mammal. The present sequence represents the human alpha-myosin heavy chain (MHC) promoter sequence, which can be used in CHF recombinant expression vector production

Sequence 2366 BP; 543 A; 657 G; 645 G; 519 T; 0 U; 2 Other;

Query Match 76.8%; Score 19.2; DB 5; Length 2366;
Best Local Similarity 87.5%; Pred. No. 5.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACATGATTGTGACA 24
Db 673 AGACAAAGAACATGATTGTGACA 650

RESULT 50

AAAS2156/C
ID AAAS2156 standard; DNA; 8050 BP.

AAAS2156;

04-DEC-2000 (first entry)

A. thaliana genomic clone of unknown function.

1-deoxy-D-xylose-5-phosphate reductoisomerase; peppermint; isoprenoid; synthesis; metabolism; chlorophyll; terpenoid; insecticidal; aroma; adhesive; ink; polymer; cytosolic; antiparasitic; plant; ds.

Arabidopsis thaliana.

WO200046346-A1.

10-AUG-2000.

27-JAN-2000; 2000WO-US002185.

03-FEB-1999; 99US-0118349P.

(UNIW) UNIV WASHINGTON STATE RES FOUND.

Croteau RB, Lange BM;

WPI; 2000-524417/47.

Nucleic acid encoding 1-deoxy-D-xylose-5-phosphate reductoisomerase, useful for enhancing or altering isoprenoid synthesis and metabolism in plants.

Example 4; Page 56-61; 65pp; English.

This is the negative (non-coding strand) of an Arabidopsis genomic clone of unknown fragment, the exonic portions of which show homology to Mentha piperita (peppermint) 1-deoxy-D-xylose-5-phosphate reductoisomerase cDNA. The peppermint reductoisomerase can be used to transform plants to enhance or alter their isoprenoid synthesis and metabolism. The enzyme catalyzes the first committed step in the conversion of 1-deoxy-D-xylose-5-phosphate to isopentenyl diphosphate, which in turn, is converted to a variety of molecules such as carotenoids, the prenyl side chains of chlorophyll, plastoquinone and tocopherols. The enzyme can therefore be used to enhance the production of chlorophyll, terpenoids, phytoalexins, toxins and deterrent compounds to improve defence against pathogens, insects and other herbivores, enhance the production of monoterpene flavour and aroma profiles, prepare synthetic intermediates for industrial use such as the synthesis of adhesives, inks and polymers, improve the yield of natural pigments extracted from plants for medicinal or culinary uses, and to enhance the yield of compounds having anti-cancer properties such as vitamin A and vitamin E

Sequence 8050 BP; 2673 A; 1517 C; 1387 G; 2473 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 3; Length 8050;
Best Local Similarity 87.5%; Pred. No. 5.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACATGATTGTGACA 24
Db 5603 AAAAAAAAAACATGATTGTGAAA 5580

Search completed: December 14, 2005, 02:42:57
Job time : 211.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:34:03 ; Search time 1752.1 Seconds
(without alignments)
667.586 Million cell updates/sec

Title: US-10-681-773-6

Perfect score: 25
Sequence: 1 aaaaaaaaaacatgattcgtgcacac 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_hnc: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_est7: *
9: gb_gss81: *
10: gb_gss82: *
11: gb_gss83: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25	100.0	933	5	BX456744 BX456744
C 2	25	100.0	1468	4	CR625326 CR625326
C 3	23.4	93.6	537	6	CA397104 CA397104
C 4	21.4	85.6	562	9	AQ513411 AQ513411
C 5	21.4	85.6	532	9	CE006606 CE006606
C 6	20.8	83.2	457	8	CX535997 CX535997
C 7	20.8	83.2	490	9	AQ676439 AQ676439
C 8	20.8	83.2	461	8	CX535992 CX535992
C 9	20.8	83.2	585	7	CV294712 CV294712
C 10	20.8	83.2	603	10	CE662963 CE662963
C 11	20.8	83.2	607	9	AZ107238 AZ107238
C 12	20.8	83.2	630	11	CR491349 CR491349
C 13	20.8	83.2	729	10	CR491349 RPCI-23-2
C 14	20.8	83.2	767	10	AX173024 BX173024
C 15	20.4	81.6	308	3	AG528328 AG528328
C 16	20.4	81.6	760	7	BP717188 BP717188
C 17	20.4	81.6	765	7	CR829278 CR829278
C 18	20.4	81.6	773	7	CR829286 CR829286
C 19	20.4	81.6	773	7	CK397627 CK397627
C 20	20.4	81.6	840	7	CK397681 CK397681
C 21	20.4	81.6	842	7	CR832404 CR832404
C 22	20.4	81.6	852	7	CR839033 CR839033

C 23	20.2	80.8	266	10	CE847912
C 24	20.2	80.8	317	6	CF772307
C 25	20.2	80.8	485	6	CX041851
C 26	20.2	80.8	486	9	AZ468621
C 27	20.2	80.8	579	10	CR755031
C 28	20.2	80.8	589	10	CE799942
C 29	20.2	80.8	605	7	COS51563
C 30	20.2	80.8	613	5	B259890
C 31	20.2	80.8	613	11	CR052145
C 32	20.2	80.8	636	11	CR108090
C 33	20.2	80.8	647	11	CR108090
C 34	20.2	80.8	653	10	CL820341
C 35	20.2	80.8	674	7	COS42928
C 36	20.2	80.8	681	1	AV676410
C 37	20.2	80.8	685	10	AG157022
C 38	20.2	80.8	690	10	C2829449
C 39	20.2	80.8	700	10	CW220895
C 40	20.2	80.8	712	10	CL185524
C 41	20.2	80.8	715	10	BX191625
C 42	20.2	80.8	715	10	CL185523
C 43	20.2	80.8	729	7	COS44112
C 44	20.2	80.8	733	7	COS49821
C 45	20.2	80.8	750	10	CW396502
C 46	20.2	80.8	751	10	CR396503
C 47	20.2	80.8	776	9	B253041
C 48	20.2	80.8	784	9	CC918461
C 49	20.2	80.8	794	9	CC918461
C 50	20.2	80.8	814	7	COS45339
C 51	20.2	80.8	815	6	CF217861
C 52	20.2	80.8	821	5	BU305387
C 53	20.2	80.8	822	7	COS51605
C 54	20.2	80.8	847	7	COS51612
C 55	20.2	80.8	854	10	CL762865
C 56	20.2	80.8	858	1	AJ814177
C 57	20.2	80.8	936	7	COS42934
C 58	20.2	80.8	943	7	COS44483
C 59	20.2	80.8	944	7	COS47156
C 60	20.2	80.8	945	7	COS50513
C 61	20.2	80.8	946	7	COS42986
C 62	20.2	80.8	950	7	COS49344
C 63	20.2	80.8	950	7	COS51101
C 64	20.2	80.8	951	7	COS44904
C 65	20.2	80.8	951	7	COS50847
C 66	20.2	80.8	955	7	COS49559
C 67	20.2	80.8	957	7	COS48352
C 68	20.2	80.8	971	7	COS47444
C 69	20.2	80.8	975	8	DN572521
C 70	20.2	80.8	977	7	COS50845
C 71	20.2	80.8	985	7	COS49311
C 72	20.2	80.8	988	7	COS51159
C 73	20.2	80.8	1046	7	COS50667
C 74	20.2	80.8	1048	9	CC270450
C 75	20.2	80.8	1053	7	COS43325
C 76	20.2	80.8	1179	9	CC227460
C 77	20.2	80.8	1183	8	DN809579
C 78	20.2	80.8	1204	8	DN807611
C 79	20.2	80.8	1204	8	DN807611
C 80	20.2	80.8	1204	8	DN807611
C 81	20.2	80.8	1204	8	DN807611
C 82	20.2	80.8	1204	8	DN807611
C 83	20.2	80.8	1204	8	DN807611
C 84	20.2	80.8	1204	8	DN807611
C 85	20.2	80.8	1204	8	DN807611
C 86	20.2	80.8	1204	8	DN807611
C 87	20.2	80.8	1204	8	DN807611
C 88	20.2	80.8	1204	8	DN807611
C 89	20.2	80.8	1204	8	DN807611
C 90	20.2	80.8	1204	8	DN807611
C 91	20.2	80.8	1204	8	DN807611
C 92	20.2	80.8	1204	8	DN807611
C 93	20.2	80.8	1204	8	DN807611
C 94	20.2	80.8	1204	8	DN807611
C 95	20.2	80.8	1204	8	DN807611

C	96	19.8	79.2	580	7	COL11408	COL11408 GR_Bb002
	97	19.8	79.2	606	9	BH528011	BH528011 BOHPF29TR
	98	19.8	79.2	629	9	BZ952006	BZ952006 CH240_44H
	99	19.8	79.2	666	9	BH248842	BH248842 BOGAJ30TR
	100	19.8	79.2	666	9	BH536629	BH536629 BOHMF47TR
	101	19.8	79.2	669	10	CG033547	CG033547 PUF646TR
	102	19.8	79.2	704	10	CLS37541	CLS37541 OB_Ba004
	103	19.8	79.2	709	9	BH937430	BH937430 Odg54h10.
C	104	19.8	79.2	725	9	BH934829	BH934829 Oe11903.
	105	19.8	79.2	732	9	BZ469447	BZ469447 BONDJ53TR
	106	19.8	79.2	734	7	CV748390	CV748390 SLL_35_00
	107	19.8	79.2	750	9	BZ650149	BZ650149 OCAPB87TC
	108	19.8	79.2	758	9	BZ034615	BZ034615 Oe108107.
	109	19.8	79.2	761	9	BH476681	BH476681 BOGX90TR
	110	19.8	79.2	770	9	BH250519	BH250519 BOGAM20TF
	111	19.8	79.2	778	9	BZ801288	BZ801288 PUFHLO7TB
	112	19.8	79.2	802	9	BZ141967	BZ141967 CH230-257
	113	19.8	79.2	802	10	DU089587	DU089587 30576_T0M
	114	19.8	79.2	818	10	CG034931	CG034931 PUFMS07TB
	115	19.8	79.2	843	9	BH426102	BH426102 BOHEC21TR
C	116	19.8	79.2	869	9	BZ796404	BZ796404 PUFEL65TB
	117	19.8	79.2	874	5	BUI33141	BUI33141 603121994
	118	19.8	79.2	880	9	BZ796412	BZ796412 PUFBL65TD
	119	19.8	79.2	913	10	CG199659	CG199659 PUFGR26TD
	120	19.8	79.2	927	4	AY809108	AY809108 Schilecso
	121	19.8	79.2	939	7	CV673845	CV673845 RET750_23
	122	19.8	79.2	960	10	CG814021	CG814021 SOYAO28TR
	123	19.8	79.2	964	9	BZ464423	BZ464423 BONKP03TR
	124	19.6	78.4	1101	10	CNS00FMH	AL070811 Drosophila
C	125	19.4	77.6	139	1	AL888911	AL888911 AL888911
	126	19.4	77.6	315	2	BF356982	BF356982 CM4-HT092
	127	19.4	77.6	325	1	AA984661	AA984661 am87909.s
	128	19.4	77.6	339	1	AA429399	AA429399 zw33906.s
	129	19.4	77.6	333	7	CO552470	CO552470 Aclly581.s
C	130	19.4	77.6	356	1	AA399296	AA399296 zt53906.s
	131	19.4	77.6	377	1	AI732746	AI732746 zw33906.x
C	132	19.4	77.6	385	1	AW513638	AW513638 x047h11.x
C	133	19.4	77.6	332	1	AI220540	AI220540 q949c10.x
C	134	19.4	77.6	333	1	AA443528	AA443528 zw33906.r
C	135	19.4	77.6	336	1	AI733921	AI733921 zt53906.y
C	136	19.4	77.6	420	8	CX400943	CX400943 JGI_XZT47
C	137	19.4	77.6	421	5	BX094228	BX094228 BX094228
C	138	19.4	77.6	443	8	R72495	R72495 yj90a09.r1
C	139	19.4	77.6	480	9	AQ126581	AQ126581 HS_3048_A
C	140	19.4	77.6	497	11	CNS07F78	AL680102 Anopheles
C	141	19.4	77.6	500	5	BX723239	BX723239 BX723239
C	142	19.4	77.6	553	1	AUI52321	AUI52321 AUI52321
C	143	19.4	77.6	565	6	CD674247	CD674247 fa08d05.y
C	144	19.4	77.6	586	9	CC061278	CC061278 MGOO_CH25
C	145	19.4	77.6	594	2	BG023139	BG023139 ddb12803.
C	146	19.4	77.6	596	7	CN073619	CN073619 EC0CBA003
C	147	19.4	77.6	642	8	CX472010	CX472010 JGI_XZG53
C	148	19.4	77.6	645	10	CE378334	CE378334 JGIr_g88s-
C	149	19.4	77.6	646	8	CX507316	CX507316 JGI_XZG52
C	150	19.4	77.6	657	8	CX367658	CX367658 JGI_XZT30

ALIGNMENTS

RESULT 1
BX456744/c
LOCUS
DEFINITION
BX456744 Homo sapiens THYMUS
5-PRIME, mRNA sequence.
ACCESSION
BX456744
VERSION
BX456744.2
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
On May 22, 2003 this sequence version replaced gi:11030820.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 9350.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0CAP003YD18-9350.r.

FEATURES
source
1..933
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP003YD18"
/issue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched, into
double-strand cDNA was digested with NotI and EcoRV sites of the
pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN

Query Match 100.0%; Score 25; DB 5; Length 933;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
DB 158 AAAAAAAAAAGCATGTTGTGACAC 134

RESULT 2
CR625326/c
LOCUS
DEFINITION
CR625326 1468 bp mRNA linear HTC 21-JUL-2004
(human).
ACCESSION
CR625326
VERSION
CR625326.1
KEYWORDS
HTC; CNSUT_cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
REFERENCE
1 (bases 1 to 1468)
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
Contact : Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 1468)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and EcoRV
sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
Location/Qualifiers

```

SOURCE
1. 1468
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP003YD18"
/tissue_type="Thymus"
/plasmid="PCWVSORT_6"

ORIGIN
Query Match      100.0%; Score 25; DB 4; Length 1468;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGCATGATTGTGACAC 25
|||||
158 AAAAAAAAAAGCATGATTGTGACAC 134

RESULT 3
CA397104 537 bp mRNA linear EST 06-NOV-2002
LOCUS CA397104.1
DEFINITION CA397104.1 Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unambiguated): cs Homo sapiens cDNA clone c86c01
5', mRNA sequence.
ACCESSION CA397104.1 GI:24734101
VERSION CA397104.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 537)
Wistow,G., Bernstein,S.L., Wyatt,M.K., Parria,R.N., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human RPE/choroid for the
NIH Bank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
12107410
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: Graeme@helix.nih.gov
Plate: 86 row: C column: 01
Seq primer: M13R1 reverse primer (ABI).
Location/Qualifiers
1. 537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="c86c01"
/tissue_type="RPE/choroid"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unambiguated): cs"
/note="Organ: Eye; Vector: PCWVSORT6; Two different donor
eyes (75-80 years old) yielded approximately 600 mg of
dissected RPE/choroid tissue. This in turn yielded 340 ug
of total RNA and 7 ug of mRNA. A directionally cloned cDNA
library in the PCWVSORT6 vector was constructed at Life
Technologies (Rockville, MD; now part of Invitrogen Corp),
essentially following the protocols of the Superscript
plasmid System (Invitrogen Corp).
<http://www.invitrogen.com/>. The library code
designation was cs. For this library, cDNA inserts were
cloned into the NotI/Mui sites of the vector. EST
analysis was performed on the unambiguated library at the
NIH Intramural Sequencing Center (NISC)."

```

```

ORIGIN
Query Match      93.6%; Score 23.4; DB 6; Length 537;
Best Local Similarity 96.0%; Pred. No. 1,1e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGCATGATTGTGACAC 25
|||||
113 AAAAAAAAAAGCATGACTGTGACAC 89

RESULT 4
A0513411 562 bp DNA linear GSS 05-MAY-1999
LOCUS A0513411
DEFINITION HS 5144 Al C11 SP66 RPCI-11 Human Male BAC library Homo sapiens
genomic clone Plate=720 Col=21 Row=E, genomic survey sequence.
ACCESSION A0513411
VERSION A0513411.1 GI:4745702
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 562)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(piet@redjond.med.buitalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buitalo.edu/ordering/bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 720 row: E column: 21
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 562.
Location/Qualifiers
1. 562
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=720 Col=21 Row=E"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI MethyIase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

```

```

ORIGIN
Query Match      85.6%; Score 21.4; DB 9; Length 562;
Best Local Similarity 95.7%; Pred. No. 6.7e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGCATGATTGTGAC 23
|||||
45 AAAAAAAAAAGCATGATTGTGAC 67

```

RESULT 5
CE006606/c 592 bp DNA linear GSS 24-SEP-2003
LOCUS tigr-gss-dog-17000320917878 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE006606
VERSION CE006606.1 GI:35004623
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 592)
Kirkness,E.F., Bathe,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Deicher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
CONTACT: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: shotgun.
Location/Qualifiers
1..592
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: Bactx1; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 85.6%; Score 21.4; DB 9; Length 592;
Best Local Similarity 95.7%; Pred. No. 6.7e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGCATGATTGTGACA 24
|||||
Db 254 AAAAAAAAAATCATGATTGTGACA 232
|||||

RESULT 6
CX535997/c 457 bp mRNA linear EST 11-JAN-2005
LOCUS TUSALVW6u769 Subtracted cDNA library of WSSV-injected shrimp
DEFINITION Litopenaeus vannamei cDNA, mRNA sequence.
ACCESSION CX535997
VERSION CX535997.1 GI:57504861
KEYWORDS EST.
SOURCE Litopenaeus vannamei (Pacific white shrimp)
ORGANISM Litopenaeus vannamei
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
Penaeidae; Litopenaeus.
1 (bases 1 to 457)
Alcivar-Warren,A., Song,L., Meehan-Meola,D., Xu,Z., Poulos,B.,
Lightner,D., Warren,W. and Xiang,J.H.
Characterization and linkage mapping of expressed sequence tags
from a subtracted cDNA library of Pacific whiteleg shrimp,
Litopenaeus vannamei, injected with white spot syndrome virus
Unpublished (2005)
Contact: Alcivar-Warren, A.
Department of Environmental and Population Health
Tufts University School of Veterinary Medicine
200 Westboro Road, North Grafton, MA 01536, USA
Tel: (508) 839-7970

JOURNAL
COMMENT

FEATURES
source
1..457
/organism="Litopenaeus vannamei"
/mol_type="mRNA"
/db_xref="taxon:6689"
/sex="not determined"
/tissue_type="most tissue types"
/cell_type="most cell types"
/dev_stage="juveniles (-3.1 grams)"
/clone_lib="Subtracted cDNA library of WSSV-injected
shrimp"
/note="Organ: whole juvenile shrimp; Vector: PCR-TRAP(x);
The whole bodies (cephalothorax and tail) of six juvenile
shrimp injected with the China strain of white spot
syndrome virus and six control (uninjected) siblings were
used for RNA isolation. The Clontech PCR-select™ cDNA
subtraction kit was used for library construction. PCR
products were ligated to PCR-TRAP(x) vector and
transformed in GH-competent cells. Lgh and Rgh primers
were used for PCR reactions."

ORIGIN
Query Match 83.2%; Score 20.8; DB 8; Length 457;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
|||||
Db 316 AAAAAAAAAATCATGATTGTGACA 293
|||||

RESULT 7
AO676439 461 bp DNA linear GSS 24-JUN-1999
LOCUS HS_2145_B2_H05_77C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone plate=2145 Col=10 Row=P, genomic survey
sequence.
ACCESSION AO676439
VERSION AO676439.1 GI:5209185
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 461)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10449764
CONTACT: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2145 Row: P Column: 10
Seq primer: T7
Class: BAC ends
High quality sequence stop: 461.
Location/Qualifiers
1..461
/organism="Homo sapiens"

REFERENCE
AUTHORS
JOURNAL
COMMENT

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2145 Col=10 Row=E"
/sex="male"
/clone_id="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelBAC11; BAC clones in E-Coli DH10B"

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 461;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGACA 24
Db 97 AATAAATTAAGCATGATTGACA 120

RESULT 8 490 bp mRNA linear EST 11-JAN-2005
CX535992/c T0ASLVWSU734 Subtracted cDNA library of WSSV-injected shrimp
LOCUS Litopenaeus vannamei cDNA, mRNA sequence.
ACCESSION CX535992
VERSION CX535992.1 GI:57504856
KEYWORDS EST.
SOURCE Litopenaeus vannamei (Pacific white shrimp)
ORGANISM Litopenaeus vannamei
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
Penaeidae; Litopenaeus.

REFERENCE

1 (bases 1 to 490)
Alciivar-Warren, A., Song, L., Meenan-Meola, D., Xu, Z., Poulos, B.,
Lightner, D., Warren, W. and Xiang, J. H.
Characterization and linkage mapping of expressed sequence tags
from a subtracted cDNA library of Pacific whiteleg shrimp,
Litopenaeus vannamei, infected with white spot syndrome virus
unpublished (2005)

TITLE

JOURNAL
COMMENT Contact: Alciivar-Warren, A.
Department of Environmental and Population Health
Tufts University School of Veterinary Medicine
200 Westboro Road, North Grafton, MA 01536, USA
Tel: (508) 839-7970
Fax: (508) 839-7091
Email: acacia.warren@tufts.edu
Seq primer: Lseq
High quality sequence stop: 490.
Location/Qualifiers
1. 490

FEATURES

source

/organism="Litopenaeus vannamei"
/mol_type="mRNA"
/db_xref="taxon:6689"
/sex="not determined"
/tissue_type="most tissue types"
/cell_type="most cell types"
/dev_stage="juveniles (~3.1 grams)"
/clone_id="Subtracted cDNA library of WSSV-injected shrimp"
/note="Organ: whole juvenile shrimp; Vector: PCR-TRAP(r);
The whole bodies (cephalothorax and tail) of six juvenile
shrimp injected with the China strain of white spot
syndrome virus and six control (uninjected) siblings were
used for RNA isolation. The CLONTECH PCR-Selectm cDNA
subtraction kit was used for library construction. PCR
products were ligated to PCR-TRAP(r) vector and
transformed in GH-competent cells. Lgh and Rgh primers
were used for PCR reactions."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 490;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGACA 24
Db 322 AAAAAAAAAAGCATTAATTGAGACA 299

RESULT 9 585 bp mRNA linear EST 23-SEP-2004
LOCUS CV294712
DEFINITION EST863089 petunia floral post-ethylene cDNA library Petunia x
hybrida cDNA clone Petunia-C2H4-23-E06 5' end, mRNA sequence.
ACCESSION CV294712
VERSION CV294712.1 GI:52584274
KEYWORDS EST.
SOURCE Petunia x hybrida
Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamiales; Solanales; Solanaceae; Petunia.
1 (bases 1 to 585)
Shibuya, K., Underwood, B., Loucas, H., Farmerie, W., Jones, M. and
Clark, D.
Petunia x hybrida EST collection
unpublished (2004)
Contact: David Clark
UF Horticulture Biotechnology Lab
University of Florida
Environmental Horticulture Department, 1545 Fifield Hall, Box
110670, Gainesville, FL 32611-0670, USA
Tel: 352-392-1831 x370
Fax: 352-392-3870
Email: dclark@mail.ifas.ufl.edu
Contact Dr. Clark (dclark@mail.ifas.ufl.edu) for clone information
Seq primer: T3 primer.
Location/Qualifiers
1. 585

FEATURES

source

/organism="Petunia x hybrida"
/mol_type="mRNA"
/cultivar="Mitchell diploid (aka. Mitchell, aka W15 in Europe)"
/db_xref="taxon:4102"
/clone="Petunia-C2H4-23-E06"
/tissue_type="all floral organs"
/lab_host="lambda ZAPri unidirectional"
/clone_id="Petunia floral post-ethylene cDNA library"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; supplier: Petunia x hybrida cv. Mitchell diploid
plants were grown from seeds to a fully flowering stage
under standard greenhouse conditions. Flowers at anthesis
stage were excised from plants grown in standard
greenhouses and treated with 2ppm exogenous ethylene.
Entire flowers were collected at 0, 5, 10, 15, 24, and 36
hours after treatment. Total RNA was extracted from each
sample, and 100 micrograms of each sample was combined for
subsequent poly A+ mRNA selection and cDNA synthesis."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 585;
Best Local Similarity 88.0%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGACAC 25
Db 70 AAAAAAAAAAGCATGATTGACAC 94

RESULT 10

CE662963/c

LOCUS CE662963 603 bp DNA linear GSS 29-SEP-2003
DEFINITION tigr-gss-dog-17000313664896 Dog library Canis familiaris genomic.
ACCESSION CE662963
VERSION CE662963.1 GI:36981831
KEYWORDS GSS.

SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1 (bases 1 to 603)
AUTHORS Kirkness, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K., Ruesch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PubMed 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: Shotgun.
FEATURES
source Location/Qualifiers
1..603
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

Query Match 83.2%; Score 20.8; DB 10; Length 603;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 559 AAAAAAAAAAGCATGATTGTGACA 536

RESULT 11
LOCUS AZ107238 607 bp DNA linear GSS 09-MAY-2000
DEFINITION RPCI-23-24P4.TV RPCI-23 Mus musculus genomic clone RPCI-23-24P4,
genomic survey sequence.
ACCESSION AZ107238
VERSION AZ107238.1 GI:7760294
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 607)
Zhao, S., Niernman, W., Feldblyum, T., Malek, J., Shatsman, S.,
Akimov, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de
Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-24P4.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.bu@iag.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.bu@iag.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tcdb/bac_ends/mouse/bac_end_intro.html

Plate: 24 row: P column: 4
Seq primer: 17
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..607
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-24P4"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

Query Match 83.2%; Score 20.8; DB 9; Length 607;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 578 AAAAAAAAAAGCATGATTGTGACA 555

RESULT 12
LOCUS CR491349 630 bp DNA linear GSS 17-NOV-2004
DEFINITION mth2-164P5F.M1 BAC end, cultivar Jemalong A17 of Medicago
truncatula, genomic survey sequence.
ACCESSION CR491349
VERSION CR491349.1 GI:48652925
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 630)
Genoscope.
Direct Submission
Submitted (10-JUN-2004) Genoscope - Centre National de Sequencage;
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)
- Web: www.genoscope.cns.fr
Location/Qualifiers
1..630
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="Jemalong A17"
/db_xref="taxon:3880"
/clone_lib="MTH2"
/note="Vector: pBelBAC11; Site_1: HindIII; Site_2:
HindIII; Cook, D.R. and Kim, D.J.
Genoscope sequence ID: mth2-164P5F.M1"

ORIGIN
Query Match 83.2%; Score 20.8; DB 11; Length 630;
Best Local Similarity 91.7%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
DB 486 AAAAAAAAAAGCATGATTGTGACA 463

RESULT 13

BX173024/c
 LOCUS BX173024 729 bp DNA linear GSS 13-MAR-2003
 DEFINITION Danio rerio genomic clone DKEX-17905, genomic survey sequence.
 ACCESSION BX173024
 VERSION BX173024.1 GI:28004729
 KEYWORDS GSS.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 729)
 Humphray, S.J., Huckle, E. and Durham, J.L.
 Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished
 This sequence was generated from the T7 end of BAC 17905. 17905 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
 Location/Qualifiers
 1..729
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEX-17905"
 /tissue_type="Testis"
 /note="vector pindigoBAC-536"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 10; Length 729;
 Best Local Similarity 91.7%; Pred. No. 1.2e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
 447 AAAAAAAAAAGCATGATTGTGACA 424

RESULT 14
 AG528328 767 bp DNA linear GSS 23-DEC-2004
 LOCUS AG528328/c
 DEFINITION Mus musculus molossinus DNA, clone:MSG01-437105.T7, genomic survey sequence.
 ACCESSION AG528328
 VERSION AG528328.1 GI:48235741
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus (Japanese wild mouse)
 ORGANISM Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaki, K. and Shirotani, T.
 Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
 Genome Res. 14 (12), 2439-2447 (2004)
 15574823
 2 (bases 1 to 767)
 Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 Direct Submission
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: hattori@gsc.riken.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the mouse BAC library MSG01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
 Tsukuba Institute, Bio Resource Center.

COMMENT

The Institute of Physical and Chemical Research (RIKEN) 3-1-1
 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp
 PRIMERS
 Sequencing : T7
 LIBRARY
 Vector : pBACe3.6
 R Site 1 : EORI
 R Site 2 : EORI.
 Location/Qualifiers
 1..767
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSG01-437105.T7"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSG01 Mouse Male BAC Library"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 10; Length 767;
 Best Local Similarity 91.7%; Pred. No. 1.2e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACA 24
 484 AAAAAAAAAAGCATGATTGTGACA 461

RESULT 15
 BP717188 308 bp mRNA linear EST 19-JUL-2004
 LOCUS BP717188/c
 DEFINITION BP717188 Osada Taira anterior neuroectoderm (ANE) PCS105 cDNA library Xenopus laevis cDNA clone XL435d17ex 3', mRNA sequence.
 ACCESSION BP717188
 VERSION BP717188.1 GI:46065781
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Xenopus.
 1 (bases 1 to 308)
 Osada, S., Kitayama, A., Ueno, N. and Taira, M.
 Expression analysis of genes which are expressed in the anterior neuroectoderm of Xenopus embryos
 Unpublished (2004)
 Contact: Masanori Taira
 Department of Biological Sciences
 Graduate School of Science, University of Tokyo; CREST, Japan
 Science and Technology Corporation, Japan
 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
 Tel: 81-03-5841-4434
 Fax: 81-03-5841-4434
 Email: m.taira@biol.s.u-tokyo.ac.jp,
 URL: <http://www.shigen.nig.ac.jp/nbrp/xenopus/est/>.
 Location/Qualifiers
 1..308
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XL435d17ex"
 /tissue_type="anterior neuroectoderm"
 /dev_stage="late gastrula (stage 12.5)"
 /clone_lib="Osada Taira anterior neuroectoderm (ANE) PCS105 cDNA library"

ORIGIN
 Query Match 81.6%; Score 20.4; DB 3; Length 308;
 Best Local Similarity 95.5%; Pred. No. 1.6e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGTTGTGA 22
 |||||
 DB 198 AAAAAAAAAATCATGTTGTGA 177

RESULT 16
 CN829278 760 bp mRNA linear EST 14-MAY-2005
 LOCUS EL3400F Brassica embryo library (EL) Brassica napus cDNA clone
 DEFINITION CN829278
 EL3400 3', mRNA sequence.
 ACCESSION CN829278
 VERSION CN829278.1 GI:65299064
 KEYWORDS EST.
 SOURCE Brassica napus (rape)
 ORGANISM Brassica napus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 760)
 AUTHORS Sharpe,A.G., Gjelvåg,B., Durkin,J. and Lydiate,D.J.
 TITLE Brassica napus ESTs
 JOURNAL Unpublished (2004)
 COMMENT Contact: Sharpe, A.G.
 Molecular Genetics
 Agriculture & Agri-Food Canada
 107 Science Place, Saskatoon, Saskatchewan, Canada, S7N0X2
 Tel: 306 956 7271
 Fax: 306 956 7247
 Email: sharpe@agr.gc.ca
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..760
 /organism="Brassica napus"
 /mol_type="mRNA"
 /cultiVar="DH12075 (double haploid line from Cresor x
 Westar cross)"
 /db_xref="taxon:3708"
 /clone="EL3400"
 /dev_stage="Mid to late embryos (4-6 mg)"
 /lab_host="E. coli Electromax DH5 alpha-e (Invitrogen)"
 /clone_lib="Brassica embryo library (EL)"
 /note="Organ: Embryos without seed coat; Vector: pSPORT1
 (modified: GCGGCCGCC*GACTATGAGCTC*cgagcgcggtcGAC);
 Site_1: NotI; Site_2: SalI; Seeds were collected by Dr.
 Francois Quellier when they were still very green (mid to
 large stage, cotyledons were formed). The seed coats were
 removed and the remaining tissue was used for cDNA library
 construction. mRNA was poly-A primed using Superscript
 Plasmid System cDNA Synthesis and Cloning kit (Invitrogen)
 After initial screening, the most abundant redundant
 clones were screened out using 22 oligos designed to match
 napins (including albumins), cruciferins, oleosins,
 trypsin inhibitor 2, cytosolic GAPDH, cyclophilins, HSP70,
 deaturase, and CAB (LHCP)."

ORIGIN
 Query Match 81.6%; Score 20.4; DB 7; Length 760;
 Best Local Similarity 95.5%; Pred. No. 1.7e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGCATGTTGTGA 24
 |||||
 DB 54 AAAAAAAAAAGATGTTGTGA 33

RESULT 17
 CN829286 765 bp mRNA linear EST 14-MAY-2005
 LOCUS EL3405F Brassica embryo library (EL) Brassica napus cDNA clone
 DEFINITION CN829286
 EL3405 3', mRNA sequence.
 ACCESSION CN829286
 VERSION CN829286.1 GI:65299072

KEYWORDS EST.
 SOURCE Brassica napus (rape)
 ORGANISM Brassica napus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 765)
 AUTHORS Sharpe,A.G., Gjelvåg,B., Durkin,J. and Lydiate,D.J.
 TITLE Brassica napus ESTs
 JOURNAL Unpublished (2004)
 COMMENT Contact: Sharpe, A.G.
 Molecular Genetics
 Agriculture & Agri-Food Canada
 107 Science Place, Saskatoon, Saskatchewan, Canada, S7N0X2
 Tel: 306 956 7271
 Fax: 306 956 7247
 Email: sharpe@agr.gc.ca
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..765
 /organism="Brassica napus"
 /mol_type="mRNA"
 /cultiVar="DH12075 (double haploid line from Cresor x
 Westar cross)"
 /db_xref="taxon:3708"
 /clone="EL3405"
 /dev_stage="Mid to late embryos (4-6 mg)"
 /lab_host="E. coli Electromax DH5 alpha-e (Invitrogen)"
 /clone_lib="Brassica embryo library (EL)"
 /note="Organ: Embryos without seed coat; Vector: pSPORT1
 (modified: GCGGCCGCC*GACTATGAGCTC*cgagcgcggtcGAC);
 Site_1: NotI; Site_2: SalI; Seeds were collected by Dr.
 Francois Quellier when they were still very green (mid to
 large stage, cotyledons were formed). The seed coats were
 removed and the remaining tissue was used for cDNA library
 construction. mRNA was poly-A primed using Superscript
 Plasmid System cDNA Synthesis and Cloning kit (Invitrogen)
 After initial screening, the most abundant redundant
 clones were screened out using 22 oligos designed to match
 napins (including albumins), cruciferins, oleosins,
 trypsin inhibitor 2, cytosolic GAPDH, cyclophilins, HSP70,
 deaturase, and CAB (LHCP)."

ORIGIN
 Query Match 81.6%; Score 20.4; DB 7; Length 765;
 Best Local Similarity 95.5%; Pred. No. 1.7e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGCATGTTGTGA 24
 |||||
 DB 48 AAAAAAAAAAGATGTTGTGA 27

RESULT 18
 CK362184 773 bp mRNA linear EST 23-DEC-2003
 LOCUS AGENCOURT_17083177 NIH_ZGC_4 Danio rerio cDNA clone IMAGE:7090415
 DEFINITION 5', mRNA sequence.
 ACCESSION CK362184
 VERSION CK362184.1 GI:40328119
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 773)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-remail.nih.gov
Tissue Procurement: John Ngai, Univ of CA, Berkeley
CDNA Library Preparation: Dr. Sumio Sugano
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM14925 row: g column: 21
High quality sequence stop: 591.
Location/Qualifiers

FEATURES

source

1..773
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:7090415"
/lab_host="DH10B Tona"
/clone_1lb="NIH ZGC 4"
/note="Organ: brain/CNS; Vector: pME18S-FL3; Site_1:
DraIII; Site_2: DraIII"

ORIGIN

Query Match 81.6%; Score 20.4; DB 7; Length 773;
Best Local Similarity 95.5%; Pred. No. 1.7e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 AAAAAAAAAAGCATGATTGCA 22
|||||
661 AAAAAAAAAAGCATGATTGCA 640

Db

RESULT 19
CK397627/c
LOCUS CK397627 795 bp mRNA linear EST 02-JAN-2004
DEFINITION AGENCOURT_17357112 NIH_ZGC_4 Danio rerio CDNA clone IMAGE:7118798
5', mRNA sequence.
ACCESSION CK397627
VERSION CK397627.1 GI:40552232
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 795)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-remail.nih.gov
Tissue Procurement: John Ngai, Univ of CA, Berkeley
CDNA Library Preparation: Dr. Sumio Sugano
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM14999 row: f column: 12
High quality sequence start: 45
High quality sequence stop: 248.
Location/Qualifiers

REFERENCE 1..795
AUTHORS /organism="Danio rerio"
TITLE /mol_type="mRNA"
JOURNAL /db_xref="taxon:7955"
COMMENT /clone="IMAGE:7119000"
/lab_host="DH10B Tona"
/clone_1lb="NIH ZGC 4"
/note="Organ: brain/CNS; Vector: pME18S-FL3; Site_1:
DraIII; Site_2: DraIII"

1..795
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:7118798"
/lab_host="DH10B Tona"
/clone_1lb="NIH ZGC 4"
/note="Organ: brain/CNS; Vector: pME18S-FL3; Site_1:
DraIII; Site_2: DraIII"

FEATURES

source

1..795
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:7118798"
/lab_host="DH10B Tona"
/clone_1lb="NIH ZGC 4"
/note="Organ: brain/CNS; Vector: pME18S-FL3; Site_1:
DraIII; Site_2: DraIII"

ORIGIN DraIII; Site_2: DraIII"

Query Match 81.6%; Score 20.4; DB 7; Length 795;
Best Local Similarity 95.5%; Pred. No. 1.7e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 AAAAAAAAAAGCATGATTGCA 22
|||||
621 AAAAAAAAAAGCATGATTGCA 600

Db

RESULT 20
CK397681/c
LOCUS CK397681 840 bp mRNA linear EST 02-JAN-2004
DEFINITION AGENCOURT_17357280 NIH_ZGC_4 Danio rerio CDNA clone IMAGE:7119000
5', mRNA sequence.
ACCESSION CK397681
VERSION CK397681.1 GI:40552286
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 840)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-remail.nih.gov
Tissue Procurement: John Ngai, Univ of CA, Berkeley
CDNA Library Preparation: Dr. Sumio Sugano
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM14999 row: n column: 22
High quality sequence start: 62
High quality sequence stop: 624.
Location/Qualifiers

FEATURES

source

1..840
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:7119000"
/lab_host="DH10B Tona"
/clone_1lb="NIH ZGC 4"
/note="Organ: brain/CNS; Vector: pME18S-FL3; Site_1:
DraIII; Site_2: DraIII"

ORIGIN

Query Match 81.6%; Score 20.4; DB 7; Length 840;
Best Local Similarity 95.5%; Pred. No. 1.7e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 AAAAAAAAAAGCATGATTGCA 22
|||||
684 AAAAAAAAAAGCATGATTGCA 663

Db

RESULT 21
CN832404/c
LOCUS CN832404 842 bp mRNA linear EST 02-JUN-2004
DEFINITION AGENCOURT_25163471 NIH_ZGC_4 Danio rerio CDNA clone IMAGE:7294735
5', mRNA sequence.
ACCESSION CN832404
VERSION CN832404.1 GI:47936157
KEYWORDS EST.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

AUTHORS 1 (bases 1 to 842)

TITLE NIH-MGC http://mgc.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: John Ngai, Univ of CA, Berkeley

cDNA Library Preparation: Dr. Sumio Sugano

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov

Plate: L1AM15315 row: m column: 05

High quality sequence start: 4

High quality sequence stop: 624.

Location/Qualifiers

1..842

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="IMAGE:7294735"

/lab_host="DH10B TONA"

/clone_1lb="NIH_ZGC_4"

/note="Organ: brain/CNS; Vector: pME185-FL3; Site_1: DraIII; Site_2: DraIII"

ORIGIN

Query Match 81.6%; Score 20.4; DB 7; Length 842;

Best Local Similarity 95.5%; Pred. No. 1.7e+03;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGA 22

689 AAAAAAAAAAGCATCTGTGA 668

Db

RESULT 22

CN839033 852 bp mRNA linear EST 02-JUN-2004

LOCUS CN839033

DEFINITION AGENCOURT_25023375 NIH_ZGC_4 Danio rerio cDNA clone IMAGE:7287820

5', mRNA sequence.

ACCESSION CN839033

VERSION CN839033.1

KEYWORDS GI:47944688

SOURCE EST.

ORGANISM Danio rerio (zebrafish)

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 852)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: John Ngai, Univ of CA, Berkeley

cDNA Library Preparation: Dr. Sumio Sugano

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: L1AM15297 row: m column: 02

High quality sequence stop: 629.

Location/Qualifiers

1..852

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="IMAGE:7287820"

/lab_host="DH10B TONA"

/clone_1lb="NIH_ZGC_4"

/note="Organ: brain/CNS; Vector: pME185-FL3; Site_1: DraIII; Site_2: DraIII"

ORIGIN

Query Match 81.6%; Score 20.4; DB 7; Length 852;

Best Local Similarity 95.5%; Pred. No. 1.7e+03;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGA 22

690 AAAAAAAAAAGCATCTGTGA 669

Db

RESULT 23

CE847912 266 bp DNA linear GSS 01-OCT-2003

LOCUS CE847912

DEFINITION tigr-gss-dog-17000332952018 Dog library Canis familiaris genomic.

genomic survey sequence.

ACCESSION CE847912

VERSION CE847912.1

KEYWORDS GI:37213010

SOURCE GSS.

ORGANISM Canis familiaris (dog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.

1 (bases 1 to 266)

Kirkness, E.F., Balna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.

The dog genome: survey sequencing and comparative analysis

Science 301 (5641), 1898-1903 (2003)

14512627

CONTACT: Kirkness EF

The Institute for Genomic Research

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-0208

Fax: 301-838-0208

Email: ekirkness@tigr.org

Class: shotgun.

Location/Qualifiers

1..266

/organism="Canis familiaris"

/mol_type="genomic DNA"

/strain="Standard Poodle"

/db_xref="taxon:9615"

/clone_1lb="Dog Library"

/note="Site_1: BckII; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 266;

Best Local Similarity 88.0%; Pred. No. 1.9e+03;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGA 25

2 AAAAAAAAAAGCATGATTGTGCAC 26

Db

RESULT 24

CF772307/c
 LOCUS 317 bp mRNA linear EST 20-OCT-2003
 DEFINITION DBSFL_32_A03_g2_A010 Drought-stressed Before flowering Sorghum
 bicolour cDNA clone DBSFL_32_A03_A010 3', mRNA sequence.
 CF772307
 ACCESSION CF772307.1 GI:37758572
 VERSION EST.
 KEYWORDS Sorghum bicolor (sorghum)
 SOURCE Sorghum bicolor
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 Clade; Panicoidae; Andropogoneae; Sorghum.
 1 (bases 1 to 317)
 Cordonnier-Pratt,M.-M., Zhang,D., McCattor,K., Nguyen,H.T. and
 Pratt,L.H.
 TITLE An EST Database from Sorghum: Subtracted pre-flowering drought
 stressed leaf tissues
 JOURNAL Unpublished (2003)
 COMMENT Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@pratt@uga.edu
 library constructed at Texas Tech University by Deshui Zhang in the
 Laboratory of Dr. Henry Nguyen. Sequencing was done in the
 Laboratory for Genomics and Bioinformatics, University of Georgia.
 Sequence ends have been trimmed to exclude vector and regions below
 phred quality 16. Three-prime sequences are presented as their
 reverse complement and have been trimmed to exclude polyA.
 Seq primer: T7 (TAATACGACTCACTATAGG)
 polyA=lee.
 FEATURES
 source
 1..317
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /mol_type="mRNA"
 /culti_var="Tx7000"
 /db_xref="taxon:4556"
 /clone="DBSFL_32_A03_A010"
 /dev_stage="Pre-flowering"
 /lab_host="ElectroMax DH10B (BRL)"
 /clone_lib="Drought-stressed before flowering"
 /note="Organ: Leaf; Vector: pBluescriptSK-; Site: 1: XhoI;
 Site: 2: EcoRI; The library was prepared from polyA+ RNA
 from leaves harvested from pre-flowering, drought-stressed
 Sorghum bicolor, cv. TX7000. Double-stranded cDNA was
 cloned unidirectionally using the Unizap system from
 Stratagene. After amplification, the library was
 subdirected by re-association hybridization. Inserts can be
 excised with XhoI and EcoRI."

Query Match 80.8%; Score 20.2; DB 6; Length 317;
 Best Local Similarity 88.0%; Pred. No. 1.9e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
 Db 158 AAAAAAAAAAGCATGATTATTATAC 134

RESULT 25
 CX041851/c 485 bp mRNA linear EST 03-JAN-2005
 LOCUS 1355350 NCCOMA 10RT#3 Oncorhynchus mykiss cDNA 5', mRNA sequence.
 DEFINITION CX041851
 ACCESSION CX041851
 VERSION CX041851.1 GI:56986205
 KEYWORDS EST.
 SOURCE Oncorhynchus mykiss (rainbow trout)
 ORGANISM Oncorhynchus mykiss
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;

REFERENCE
 AUTHORS 1 (bases 1 to 485)
 TITLE Yao,J., Gahr,S. and Rexroad,C.E.
 JOURNAL 10RT#3 egg, NCCOMA/MVU EST Project, Phase II
 COMMENT Unpublished (2004)
 CONTACT: Rexroad CE
 USDA, ARS, National Center for Cool and Cold Water Aquaculture
 11876 Leetown Road, Kearneysville, WV 25430, USA
 Tel: 304 724 8340 x2129
 Fax: 304 725 0351
 Email: crexroad@nccoma.ars.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_al option. Vector identified with
 cross_match v0.990329.
 Plate: 152 row: N column: 4
 Seq primer: GTAATACGACTCACTATAGG.
 FEATURES
 source
 1..485
 Location/Qualifiers
 /organism="Oncorhynchus mykiss"
 /mol_type="mRNA"
 /db_xref="taxon:8022"
 /lab_host="DH10B"
 /clone_lib="NCCOMA 10RT#3"
 /note="Vector: PCMV Sport6.0; MVU oocyte library RT-EGG."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 8; Length 485;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
 Db 469 AAAAAAAAAATCATGATTGTCACTC 445

RESULT 26
 AZ468621/c 486 bp DNA linear GSS 04-OCT-2000
 LOCUS 1M0281E23R Mouse 10kb plasmid UGCG1M library Mus musculus genomic
 DEFINITION clone UGCG1M0281E23 R, genomic survey sequence.
 ACCESSION AZ468621
 VERSION AZ468621.1 GI:10626842
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 486)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T.,
 Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
 Niederhausern,A. and Wright,D. Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0281 row: E column: 23
 Seq primer: CAACAAGAAACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 486.
 Location/Qualifiers
 1..486
 /organism="Mus musculus"
 /mol_type="genomic DNA"

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGGCM0281E23"
/sex="Male"
/lab_host="R. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UGGCM library"
/note="Vector: PWD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114|g1473072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 486;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACAC 25
461 AAAAAAAAAACCTGATGTGACAC 437

RESULT 27
CW755031/c 579 bp DNA linear GSS 09-NOV-2004
LOCUS CG_BBA0057K24.r OG_BBA_Oryza glaberrima genomic clone OG_BBA0057K24
DEFINITION 3' genomic survey sequence.
ACCESSION CW755031
VERSION CW755031.1 GI:55593701
KEYWORDS GSS.
SOURCE Oryza glaberrima (African rice)
ORGANISM Oryza glaberrima
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.
1 (bases 1 to 579)
Kim, H., Yu, Y., Wisotzki, M., Byrne, M., Stum, D., Smart, D., Rao, K.,
Luo, M., Jetty, R., Kudrna, D., Muller, C., Hatfield, J., Soderlund, C.
and Wing, R.
OMAP

TITLE

JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: twing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0057 row: K column: 24
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.

REFERENCE
AUTHORS

FEATURES

source 1. .579
/organism="Oryza glaberrima"
/mol_type="genomic DNA"
/db_xref="taxon:4538"

/clone="OG_BBA0057K24"
/cissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OG_BBA"
/note="Vector: pAG1BAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 579;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACAC 25
560 AAAAAAAAAATCATGATGTGACAC 536

RESULT 28
CE799942 589 bp DNA linear GSS 30-SEP-2003
LOCUS tigr-gss-dcg-17000317693815 Dog Library Canis familiaris genomic.
DEFINITION genomic survey sequence.
ACCESSION CE799942
VERSION CE799942.1 GI:37140839
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 589)
Kirkness, E.F., Balina, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkne@tigr.org
Class: shotgun.
Location/Qualifiers
1. .589
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

COMMENT

PERMUTED
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkne@tigr.org
Class: shotgun.
Location/Qualifiers
1. .589
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

FEATURES
source

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 589;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGACAC 25
429 AAAAAAAAAAGTTGATTGGAAC 453

RESULT 29

COS51563/c 605 bp mRNA linear EST 01-SEP-2004
LOCUS COS51563
DEFINITION AcLY1124 See lamprey Acly Petromyzon marinus CDNA, mRNA sequence.
ACCESSION COS51563
VERSION COS51563.1 GI:51799893
KEYWORDS EST.
SOURCE Petromyzon marinus (sea lamprey)
ORGANISM Petromyzon marinus

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia; Petromyzontiformes; Petromyzontidae; Petromyzon.
1 (bases 1 to 605)
Pancer, Z., Mayer, W. E., Klein, J. and Cooper, M. D.
Prototypic T-cell receptor and CD4-like coreceptor expressed in lymphocytes of the agnathan sea lamprey
Proc. Natl. Acad. Sci. U.S.A. 101 (36), 13273-13278 (2004)
15328402
Contact: Pancer, Zeev
Division of Developmental and Clinical Immunology
The University of Alabama at Birmingham
378 Wallace Tumor Institute, 1530 Third Avenue, South, Birmingham, AL 35294-3300
Tel: 205-975-5812
Fax: 205-975-7218
Email: zpancer@uab.edu.
Location/Qualifiers
1. 605
/organism="Petromyzon marinus"
/mol_type="mRNA"
/db_xref="taxon:7757"
/cell_type="lymphocyte"
/dev_stage="immune stimulated larvae"
/note="Vector: pGEM-T Easy; lymphocyte mRNA ESTs from PCR subtracted cDNA libraries of immune stimulated larvae. All are single pass 5' or 3' sequences randomly cloned in pGEM-T Easy (Promega)."

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 605;
Best Local Similarity 88.0%; Pred. No. 2e+03; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
|||||
564 AAAAAAAAAAGCATGATTGTGCAC 540

RESULT 30
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

B0259890 613 bp mRNA linear EST 26-NOV-2002
6035011703f1 CSROCHN51 Gallus gallus cDNA clone CHEST420e1 5', mRNA
sequence.
B0259890
B0259890.1 GI:25527062
EST.
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
1 (bases 1 to 613)
Boardman, P. E., Sanz-Burguero, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 613
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hixex"
/db_xref="taxon:9031"
/clone="CHEST420e1"

FEATURES
source

ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 613;
Best Local Similarity 88.0%; Pred. No. 2e+03; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
|||||
528 AAAAAAAAAAGCATGATTGTGCAC 504

RESULT 31
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CR052145 613 bp DNA linear GSS 05-JUL-2004
Reverse strand read from insert in 3'HPT insertion targeting and chromosome engineering clone MHP41105, genomic survey sequence.
CR052145
CR052145.1 GI:49785284
GSS; genome survey sequence; MICE.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 613)
Adams, D. J., Biggs, P. J., Cox, A. V., Davies, R. M., van der Weyden, L., Jonkers, J., Smith, J. J., Plumb, R. W., Taylor, R. G., Nishijima, I., Yu, Y., Rogers, J. and Bradley, A.
Direct Submision
Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. <http://www.sanger.ac.uk/MICE>
Location/Qualifiers
1. 613
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHP41105"
/clone_1ib="MHP"

FEATURES
source

ORIGIN

Query Match 80.8%; Score 20.2; DB 11; Length 613;
Best Local Similarity 88.0%; Pred. No. 2e+03; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGCAC 25
|||||
178 AAAAAAAAAAGCATGATTGTGCAC 202

RESULT 32
LOCUS
DEFINITION
ACCESSION
VERSION

CR108090 636 bp DNA linear GSS 05-JUL-2004
Forward strand read from insert in 5'HPT insertion targeting and chromosome engineering clone MHP417n12, genomic survey sequence.
CR108090
CR108090.1 GI:49855505

KEYWORDS GSS; genome survey sequence; MICR.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 636)
Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L., Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y., Rogers,J. and Bradley,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. <http://www.sanger.ac.uk/MICR>
Location/Qualifiers
FEATURES
source
1..636
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN47n12"
/clone_1fb="MHPN"
ORIGIN
Query Match 80.8%; Score 20.2; DB 11; Length 636;
Best Local Similarity 88.0%; Pred. No. 2e+03; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3;
QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
Db 312 AAAAAAAAAATCATTTATGTGACTC 288
RESULT 33
CL820341/c 647 bp DNA linear GSS 06-JUL-2004
LOCUS Forward strand read from insert in 5'HPRT insertion targeting and chromosome engineering clone MHPN194013, genomic survey sequence.
DEFINITION CR117340
CR117340.1 GI:49864773
VERSION GSS; genome survey sequence; MICR.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 647)
Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L., Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y., Rogers,J. and Bradley,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. <http://www.sanger.ac.uk/MICR>
Location/Qualifiers
FEATURES
source
1..647
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN194013"
/clone_1fb="MHPN"
ORIGIN
Query Match 80.8%; Score 20.2; DB 11; Length 647;
Best Local Similarity 88.0%; Pred. No. 2e+03; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3;
QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
Db 296 AAAAAAAAAATCATTTATGTGACTC 272
RESULT 34
CL820341/c 653 bp DNA linear GSS 09-AUG-2004
LOCUS CL820341
DEFINITION OR_CBa0038018.r OR_CBa Oryza rufipogon genomic clone OR_CBa0038018

ACCESSION 3', genomic survey sequence.
CL820341
VERSION CL820341.1 GI:51065951
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 653)
Kim,H., Yu,Y., Matsuoka,M., Yost,D., Stum,D., Rao,K., Luo,M., Jettly,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: twing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0038 row: 0 column: 18
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
Location/Qualifiers
FEATURES
source
1..653
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0038018"
/cissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_1fb="OR_CBa"
/note="Vector: pAG1BAC1; Site_1: HindIII; Site_2: HindIII; drk treated 36 hrs before harvest"
ORIGIN
Query Match 80.8%; Score 20.2; DB 10; Length 653;
Best Local Similarity 88.0%; Pred. No. 2e+03; 3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3;
QY 1 AAAAAAAAAAGCATGTTGTGACAC 25
Db 560 AAAAAAAAAATCATGATGTGACAC 536
RESULT 35
COS42928 674 bp mRNA linear EST 01-SEP-2004
LOCUS LYEST10158 Sea lamprey lyEST Petromyzon marinus cDNA, mRNA
DEFINITION sequence.
ACCESSION COS42928
VERSION COS42928.1 GI:51787918
KEYWORDS EST.
SOURCE Petromyzon marinus (sea lamprey)
ORGANISM Petromyzon marinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia; Petromyzoniformes; Petromyzontidae; Petromyzon.
REFERENCE 1 (bases 1 to 674)
Pancer,Z., Mayer,W.B., Klein,J. and Cooper,M.D.
TITLE Prototypic T-cell receptor and CD4-like coreceptor expressed in lymphocytes of the agnathan sea lamprey
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (36), 13273-13278 (2004)
COMMENT Contact: Pancer, Zeev
Division of Developmental and Clinical Immunology
The University of Alabama at Birmingham
378 Wallace Tumor Institute, 1550 Third Avenue, South, Birmingham,

AL 35294-3300
Tel: 205-975-5812
Fax: 205-975-7218
Email: zpancer@uab.edu.
Location/Qualifiers
1. .674
/organism="Petrotyzon marinus"
/mol_type="mRNA"
/db_xref="taxon:7757"
/cell_type="lymphocyte"
/dev_stage="unstimulated larvae"
/clone_lib="Sea lamprey lyEST"
/notes="Vector: Lambda ZAP Express; lymphocyte mRNA ESTs from unstimulated larvae. All are from arrayed colonies from a directionally cloned cDNA library in Lambda ZAP Express (Stratagene). All are single pass 5' sequences."

ORIGIN
Query Match 80.8%; Score 20.2; DB 7; Length 674;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
|||||
298 AAAAAAAAAAGCATGATTGTGCTAC 274

RESULT 36
AV676410 681 bp mRNA linear EST 26-MAY-2005
LOCUS AV676410 Nori Satoh unpublished cDNA library Ciona intestinalis
DEFINITION cDNA clone rc1db4k18 3', mRNA sequence.
ACCESSION AV676410
VERSION AV676410.1 GI:10114409
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 681)
Satoh,N., Satou,Y., Kohara,Y. and Shin-I,T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: sato@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. .681
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rc1db4k18"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_lib="Nori Satoh unpublished cDNA library"

ORIGIN
Query Match 80.8%; Score 20.2; DB 1; Length 681;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
|||||
280 AAAAAAAAAAGCATGCTTGTACAC 304

RESULT 37
AG157022 685 bp DNA linear GSS 09-JAN-2002
LOCUS AG157022

DEFINITION Pan troglodytes DNA, clone: RP43-021N03.TU, genomic survey sequence.
ACCESSION AG157022
VERSION AG157022.1 GI:16686700
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Pan.
1
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of library RP43-021N03
Unpublished
2 (bases 1 to 685)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenho-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbases@gsc.riken.go.jp, URL:http://hsp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library RP43-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: TU
LIBRARY
Vector : pBACe3.6
R.site 1 : EcoRI
R.site 2 : EcoRI.
Location/Qualifiers
1. .685
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-021N03.TU"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP43-43 Chimpanzee Male BAC Library"

ORIGIN
Query Match 80.8%; Score 20.2; DB 10; Length 685;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTGTGACAC 25
|||||
532 AAAAAAAAAAGCCTGATGTGACAC 556

RESULT 38
CZ829449 690 bp DNA linear GSS 26-JULY-2005
LOCUS CZ829449/C OC_Ba0210L16.r OC_Ba Oryza coarctata genomic clone OC_Ba0210L16
DEFINITION 3', genomic survey sequence.
ACCESSION CZ829449
VERSION CZ829449.1 GI:71269302
KEYWORDS GSS.
SOURCE Oryza coarctata (Porteresia coarctata)
ORGANISM Oryza coarctata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrharioideae; Oryzaceae; Oryza.
1 (bases 1 to 690)
Kim,H., Collura,K., Wisotzki,M., Byrne,M., Stum,D., Smart,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and Wing,R.
OMAP (Oryza Map Alignment Project) - Arizona Genomics Institute
Unpublished (2005)
COMMENT Contact: Rod A. Wing

Arizona Genomics Institute

University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595
Fax: 520 621 1259

Email: twing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0210 row: L column: 16
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.

FEATURES

source Location/Qualifiers

1..650

/organism="Oryza coarctata"

/mol_type="genomic DNA"

/db_xref="taxon:77588"

/clone_lib="OC_BA0210L16"

/issue_type="leaves"

/dev_stage="mature"

/lab_host="DH10B"

/note="Vector: pGIBAC1, Site_1: HindIII, Site_2: HindIII"

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 690;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAAAAAAAAACATGATGTGACAC 25

240 AAAAAAAAAAGAACGATTGTGACAC 216

RESULT 39 700 bp DNA linear GSS 30-OCT-2004
LOCUS CM220895
DEFINITION 104.555..1197608.116.37156.084 Sorghum methylation filtered library
(LibID: 104) Sorghum bicolor genomic clone 1197608, genomic survey
sequence.

ACCESSION CM220895
VERSION CM220895.1 GI:54929722
KEYWORDS GSS.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

COMMENT Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 700)

REFERENCE Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
McMenamy,J., Smith,M., Holeman,H., Roe,B.A., Wiley,G., Korf,I.F.,
Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
Martensen,R.A.

AUTHORS Sorghum genome sequencing by methylation filtration
PLOS Biol. 3 (1), e13 (2005)

TITLE Sorghum genome sequencing by methylation filtration
JOURNAL PLOS Biol. 3 (1), e13 (2005)
PUBMED 15660154
COMMENT Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: 655 row: m column: 24
Seq primer: T3 Reverse
Class: methylation filtered
High quality sequence stop: 700.

FEATURES

source Location/Qualifiers

1..700

/organism="Sorghum bicolor"

/mol_type="genomic DNA"

/db_xref="taxon:4558"

/db_xref="taxon:4558"

/clone_lib="1197608"

/clone_lib="Sorghum methylation filtered library (LibID: 104)"

/note="Organ: leaf; Vector: pBCK(-); Site_1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBCK(-) vector
and electroporated into E. coli cells. This is a
methylation filtered library."

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 700;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAAAAAAAAACATGATGTGACAC 25

173 AAAAAAAAAACATGATGTGACAC 197

RESULT 40 712 bp DNA linear GSS 06-JAN-2004
LOCUS CL185524/c
DEFINITION 104.399.10899592.116.32397.054 Sorghum methylation-filtered library
(LibID: 104) Sorghum bicolor genomic clone 10899592, genomic survey
sequence.

ACCESSION CL185524
VERSION CL185524.1 GI:40698047
KEYWORDS GSS.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

COMMENT Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 712)

REFERENCE Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
McMenamy,J., Smith,M., Holeman,H., Roe,B.A., Wiley,G., Korf,I.F.,
Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
Martensen,R.A.

AUTHORS Sorghum genome sequencing by methylation filtration
PLOS Biol. 3 (1), e13 (2005)

TITLE Sorghum genome sequencing by methylation filtration
JOURNAL PLOS Biol. 3 (1), e13 (2005)
PUBMED 15660154
COMMENT Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: 399 row: l column: 16
Seq primer: T3 Reverse
Class: methylation filtered
High quality sequence stop: 712.

FEATURES

source Location/Qualifiers

1..712

/organism="Sorghum bicolor"

/mol_type="genomic DNA"

/db_xref="taxon:4558"

/clone_lib="10899592"

/clone_lib="Sorghum methylation-filtered library (LibID: 104)"

/note="Organ: leaf; Vector: pBCK(-); Site_1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to 5
kb fraction, ligated into HincII-digested pBCK(-) vector
and electroporated into E. coli cells. This is a
methylation-filtered library."

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 712;
Best Local Similarity 88.0%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTCGACAC 25
 Db 312 AAAAAAAAAATCATGATTCGACAC 288

RESULT 41
 LOCUS BX191625 715 bp DNA linear GSS 13-MAR-2003
 DEFINITION Danio rerio genomic clone DKEX-20208, genomic survey sequence.
 ACCESSION BX191625
 VERSION BX191625.1 GI:28023511
 KEYWORDS GSS.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 715)
 Humphray,S.J., Huckle,E. and Durham,J.L.
 Direct Submission
 Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished

COMMENT This sequence was generated from the T7 end of BAC 20208. 20208 is part of the Daniokey BAC library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_rerio/

FEATURES
 source Location/Qualifiers
 1..715
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEX-20208"
 /issue_type="Testis"
 /note="Vector pIndigoBAC-536"

ORIGIN
 Query Match 80.8%; Score 20.2; DB 10; Length 715;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTCGACAC 25
 Db 249 AAAAAAAAAATCATGATTCGACAC 273

RESULT 42
 LOCUS CL185523 715 bp DNA linear GSS 06-JAN-2004
 DEFINITION 104_399_10899592_114_32393_054 Sorghum methylation-filtered library (LibID: 104) Sorghum bicolor genomic clone 10899592, genomic survey sequence.
 ACCESSION CL185523
 VERSION CL185523
 KEYWORDS GSS.
 SOURCE Sorghum bicolor (sorghum)
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 715)
 Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D., Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K., Momeny,J., Smith,M., Holman,H., Roe,B.A., Wiley,G., Korf,I.F., Rabinowicz,P.D., Lakey,N., McComb,W.R., Jeddeloh,J.A. and Mortensen,R.A.
 Sorghum genome sequencing by methylation filtration
 PLOS Biol. 3 (1), e13 (2005)
 Contact: Bedell JA
 Orion Genomics, LLC
 4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979
 Fax: 314 615 5975
 Email: jbedell@oriongenomics.com
 Plate: 399 row: 1 column: 16
 Seq primer: M13/pUC Forward
 Clase: methylation filtered
 High quality sequence stop: 715.
 Location/Qualifiers
 1..715
 /organism="Sorghum bicolor"
 /mol_type="genomic DNA"
 /cultivar="ATx623"
 /db_xref="taxon:4558"
 /clone="10899592"
 /clone_11b="Sorghum methylation-filtered library (LibID: 104)"
 /note="Organ: leaf; Vector: pBCSK(-); Site: 1; HindII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HindII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation-filtered library."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 10; Length 715;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGCATGATTCGACAC 25
 Db 612 AAAAAAAAAATCATGATTCGACAC 636

RESULT 43
 LOCUS COS44112/c 729 bp mRNA linear EST 01-SEP-2004
 DEFINITION LYEST11818 Sea lamprey LYEST Petromyzon marinus CDNA, mRNA sequence.
 ACCESSION COS44112
 VERSION COS44112.1 GI:51790305
 KEYWORDS EST.
 SOURCE Petromyzon marinus (sea lamprey)
 ORGANISM Petromyzon marinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia; Petromyzoniformes; Petromyzontidae; Petromyzon.
 1 (bases 1 to 729)
 Pancer,Z., Mayer,W.E., Klein,J. and Cooper,M.D.
 Prototypic T-cell receptor and CD4-like coreceptor expressed in lymphocytes of the agnathan sea lamprey
 Proc. Natl. Acad. Sci. U.S.A. 101 (36), 13273-13278 (2004)
 Contact: Pancer, Zeev
 Division of Developmental and Clinical Immunology
 The University of Alabama at Birmingham
 378 Wallace Tumor Institute, 1530 Third Avenue, South, Birmingham, AL 35294-3300
 Tel: 205-975-5812
 Fax: 205-975-7218
 Email: zpancer@uab.edu

FEATURES
 source Location/Qualifiers
 1..729
 /organism="Petromyzon marinus"
 /mol_type="mRNA"
 /db_xref="taxon:7757"
 /cell_type="lymphocyte"
 /dev_stage="unstimulated larvae"
 /clone_11b="Sea lamprey LYEST"
 /note="Vector: lambda ZAP Express; lymphocyte mRNA ESTs from unstimulated larvae. All are from arrayed colonies from a directionally cloned cDNA library in lambda ZAP Express (Stratagene). All are single pass 5' sequences."

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 729;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGCAC 25
 |||||
 370 AAAAAAAAAAGCATGATTGTCTAC 346

RESULT 44
 LOCUS COS49821/c 733 bp mRNA linear EST 01-SEP-2004
 DEFINITION LYEST7547 Sea lamprey LYEST Petromyzon marinus CDNA, mRNA sequence.
 ACCESSION COS49821
 VERSION COS49821.1 GI:51798137
 KEYWORDS EST.
 SOURCE Petromyzon marinus (sea lamprey)
 ORGANISM Petromyzon marinus; Chordata; Vertebrata; Hyperoartia; Eukaryota; Metazoa; Petromyzontidae; Petromyzon.
 1 (bases 1 to 733)
 Pancer/Z., Mayer,W.E., Klein,J. and Cooper,M.D.
 Prototypic T-cell receptor and CD4-like coreceptor expressed in lymphocytes of the agnathan sea lamprey
 Proc. Natl. Acad. Sci. U.S.A. 101 (36), 13273-13278 (2004)
 15328402
 COMMENT Contact: Pancer, Zeev
 Division of Developmental and Clinical Immunology
 The University of Alabama at Birmingham
 378 Wallace Tumor Institute, 1550 Third Avenue, South, Birmingham, AL 35294-3300
 Tel: 205-975-5812
 Fax: 205-975-7218
 Email: zpancer@uab.edu.

FEATURES
 source Location/Qualifiers
 1..733
 /organism="Petromyzon marinus"
 /mol_type="mRNA"
 /db_xref="taxon:7757"
 /cell_type="lymphocyte"
 /dev_stage="unstimulated larvae"
 /clone_lib="Sea lamprey LYEST"
 /note="Vector: Lambda ZAP Express; lymphocyte mRNA ESTs from unstimulated larvae. All are from arrayed colonies from a directionally cloned cDNA library in Lambda ZAP Express (Stratagene). All are single pass 5' sequences."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 7; Length 733;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGCAC 25
 |||||
 602 AAAAAAAAAAGCATGATTGTCTAC 578

RESULT 45
 LOCUS CW396502 750 bp DNA linear GSS 01-NOV-2004
 DEFINITION fdbb001f084p13f0 Sorghum methylation filtered library (libid: 104)
 Sorghum bicolor genomic clone fdbb001f084p13, genomic survey sequence.
 ACCESSION CW396502
 VERSION CW396502.1 GI:55114946
 KEYWORDS GSS.
 SOURCE Sorghum bicolor (sorghum)
 ORGANISM Sorghum bicolor; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 750)
 Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,

Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K., McManamy,J., Smith,M., Holuman,H., Roe,B.A., Wiley,G., Korf,I.F., Rabinowicz,P.D., Lakey,N., McComble,W.R., Jeddeloh,J.A. and Martienssen,R.A.
 Sorghum genome sequencing by methylation filtration
 PLOS Biol. 3 (1), e13 (2005)
 15660154

TITLE
 JOURNAL PLOS Biol. 3 (1), e13 (2005)
 PUBMED 15660154

COMMENT Contact: Bedell JA
 Orion Genomics, LLC
 4041 Forest Park Ave, St. Louis, MO 63108, USA
 Tel: 314 615 6979
 Fax: 314 615 5975
 Email: jbedell@oriongenomics.com
 Plate: fdbb001f084 row: p column: 13
 Seq primer: f Forward
 Class: methylation filtered
 High quality sequence stop: 750.

FEATURES
 source Location/Qualifiers
 1..750
 /organism="Sorghum bicolor"
 /mol_type="genomic DNA"
 /cultivar="ATx623"
 /db_xref="taxon:4558"
 /clone_lib="fdbb001f084p13"
 /clone_lib="Sorghum methylation filtered library (libid: 104)"
 /note="Organ: leaf; Vector: pBCSK(-); Site: 1; HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 10; Length 750;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCATGATTGTGCAC 25
 |||||
 553 AAAAAAAAAATCATGATTGTGCAC 577

RESULT 46
 LOCUS CW396503/c 751 bp DNA linear GSS 01-NOV-2004
 DEFINITION fdbb001f084p13k0 Sorghum methylation filtered library (libid: 104)
 Sorghum bicolor genomic clone fdbb001f084p13, genomic survey sequence.
 ACCESSION CW396503
 VERSION CW396503.1 GI:55114947
 KEYWORDS GSS.
 SOURCE Sorghum bicolor (sorghum)
 ORGANISM Sorghum bicolor; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 751)
 Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D., Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K., McManamy,J., Smith,M., Holuman,H., Roe,B.A., Wiley,G., Korf,I.F., Rabinowicz,P.D., Lakey,N., McComble,W.R., Jeddeloh,J.A. and Martienssen,R.A.
 Sorghum genome sequencing by methylation filtration
 PLOS Biol. 3 (1), e13 (2005)
 15660154

TITLE
 JOURNAL PLOS Biol. 3 (1), e13 (2005)
 PUBMED 15660154

COMMENT Contact: Bedell JA
 Orion Genomics, LLC
 4041 Forest Park Ave, St. Louis, MO 63108, USA
 Tel: 314 615 6979
 Fax: 314 615 5975
 Email: jbedell@oriongenomics.com
 Plate: fdbb001f084 row: p column: 13

Seq primer: K Reverse
 Class: methylation filtered
 High quality sequence stop: 751.
 Location/Qualifiers

FEATURES

Source

1. 751
 /organism="Sorghum bicolor"
 /mol_type="genomic DNA"
 /cultivar="ATx623"
 /db_xref="taxon:4558"
 /clone_id="f8b001f084p13"
 /note="Sorghum methylation filtered library (Libid: 104)"
 /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 751;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTGCAC 25
 Db 499 AAAAAAAAAATCATGATTGTGCAC 475

RESULT 47
 BU227097/c 773 bp mRNA linear EST 26-NOV-2002

LOCUS BU227097 773 bp mRNA linear EST 26-NOV-2002
 DEFINITION 60394714F1 CSECHN23 Gallus gallus cDNA clone CHEST901p12 5', mRNA
 sequence.
 BU227097
 VERSION BU227097.1 GI:25464663
 KEYWORDS EST
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 773)
 Boatman,P.E., Sam-Esquero,J., Overton,I.M., Burr,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 12445392

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 PUBMED
 COMMENT
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers

FEATURES

Source

1. 773
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="white Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST901p12"
 /dev_stage="22"
 /lab_host="DH10B"
 /clone_id="CSECHN23"
 /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was bluntended, ligated to NotI adapters, digested with

EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 773;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTGCAC 25
 Db 151 AAAAAAAAAATCATGATTGTGCAC 127

RESULT 48
 BZ653041 776 bp DNA linear GSS 29-JAN-2003

LOCUS BZ653041 776 bp DNA linear GSS 29-JAN-2003
 DEFINITION OGC8230TC ZM 0.7-1.5 KB Zea mays genomic clone ZMMBMA0137F12,
 genomic survey sequence.
 BZ653041
 VERSION BZ653041.1 GI:28120916
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 776)
 Resnick,A., Fraser,C.M., Buddiman,M.A., Bedell,J.A., Rohlfing,T.,
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
 Consortium for Maize Genomics
 Unpublished (2002)
 Other GSSs: OGC8230TM
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Class: methylation filtered.
 Location/Qualifiers

FEATURES

Source

1. 776
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBMA0137F12"
 /clone_id="ZM_0.7-1.5 KB"
 /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
 methylation filtered genomic DNA library"

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 776;
 Best Local Similarity 88.0%; Pred. No. 2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGCATGATTGTGCAC 25
 Db 709 AAAAAAAAAAGCATGATTGTGCAC 733

RESULT 49
 COS51600 787 bp mRNA linear EST 01-SEP-2004
 LOCUS COS51600 787 bp mRNA linear EST 01-SEP-2004
 DEFINITION Acly1180 Sea lamprey Acly Petromyzon marinus cDNA, mRNA sequence.
 ACCESSION COS51600
 VERSION COS51600.1 GI:51799930

```

SOURCE ORGANISM EST. Petromyzon marinus (sea lamprey)
SOURCE ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
SOURCE ORGANISM Petromyzon marinus Petromyzontiformes; Petromyzontidae; Petromyzon.
REFERENCE 1 (bases 1 to 787)
AUTHORS Pancer,Z., Mayer,W.E., Klein,J. and Cooper,M.D.
TITLE Prototypic T-cell receptor and Cys-like coreceptor expressed in lymphocytes of the agnathan sea lamprey
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (36), 13273-13278 (2004)
COMMENT 15328402
PDBMED Contact: Pancer, Zeev
Division of Developmental and Clinical Immunology
The University of Alabama at Birmingham
378 Wallace tumor Institute, 1530 Third Avenue, South, Birmingham, AL 35294-3300
Tel: 205-975-5812
Fax: 205-975-7218
Email: zpancer@uab.edu.
FEATURES source location/Qualifiers
source 1..787
/organism="Petromyzon marinus"
/mol_type="mRNA"
/db_xref="taxon:7757"
/cell_type="lymphocyte"
/dev_stage="immune stimulated larvae"
/clone_lib="Sea Lamprey Acly"
/note="Vector: pGEM-T Easy; lymphocyte mRNA ESTs from PCR subtracted cDNA libraries of immune stimulated larvae. All are single pass 5' or 3' sequences randomly cloned in pGEM-T Easy (Promega)."
ORIGIN
Query Match 80.8%; Score 20.2; DB 7; Length 787;
Best Local Similarity 88.0%; Pred. No.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGCATATTGTGCAC 25
Db 320 AAAAAAAAAAGCATATTGTGTCTAC 344
RESULT 50
CC918461.c 794 bp DNA linear GSS 08-AUG-2003
LOCUS t006110ba.f1 TAMBT Bos taurus genomic clone t006110ba, genomic
DEFINITION survey sequence.
ACCESSION CC918461
VERSION CC918461.1 GI:33548960
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Peccora; Bovidae; Bovinae; Bos.
1 (bases 1 to 794)
lin,S., Najjar,F.Z., Adelson,D., Gill,C.A. and Roe,B.A.
Bovine BAC End Sequences from Library TAMBT
Unpublished (2003)
COMMENT Contact: Bruce A. Roe
Advanced Center for Genome Technology
University of Oklahoma Department of Chemistry and Biochemistry
620 Parrington Oval, Room 208, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Class: BAC ends
High quality sequence start: 29
High quality sequence stop: 621.
location/Qualifiers
1..794
/organism="Bos taurus"
/mol_type="genomic DNA"
FEATURES source

```

```

      /strain="Angus bull T A M U Shoshone Y6 11519666"  

      /db xref="taxon:9913"  

      /clone="t006110ba"  

      /sex="Male"  

      /cell_type="Blood"  

      /clone_id="TAMBTn"  

      /note="Vector: pBelOBAC11; Site 1: HindIII; Site 2:  

      HindIII; TAMBT Bovine BAC library (Male) produced by Texas  

      A&M University, Department of Animal Science."  

ORIGIN  

Query Match.          80.8%; Score 20.2; DB 9; Length 794;  

Best Local Similarity 88.0%; Pred. No. 2e+03;  

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0,  

QY      1 AAAAAAAAAAGCATGATTGTGCAC 25  

        |||||         |||||  

Db       482 AAAAAAAAAAATGATTGGGACAC 458  


```

Search completed: December 14, 2005, 07:35:12
Job time : 1759.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-7
Perfect score: 25
Sequence: 1 aaaaaaaaaagtcacatcagata 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_in: *
3: gb_env: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pr: *
9: gb_ro: *
10: gb_sbs: *
11: gb_sy: *
12: gb_un: *
13: gb_vl: *
14: gb_hcg: *
15: gb_pl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	140712	14	AC022223 Homo sapi
2	23.4	93.6	88814	8	AC010270 Homo sapi
3	23.4	93.6	103701	8	AC008434 Homo sapi
4	23.4	93.6	163303	14	AC104467 Homo sapi
5	22.4	93.6	223885	8	AY275681 Homo sapi
6	22.4	89.6	62860	8	AL162395 Human DNA
7	22.4	89.6	110000	14	AC162929_1
8	22.4	89.6	200157	9	AC140409 Mus muscu
9	21.8	87.2	5953	6	AX348364 Sequence
10	21.8	87.2	10151	6	AX347335 Sequence
11	21.8	87.2	89934	15	AC007067 Genomic s
12	21.8	87.2	99416	15	AC009398 Genomic s
13	21.8	87.2	167510	9	AC160146 Mus muscu
14	21.8	87.2	189457	9	AC135019 Mus muscu
15	21.8	87.2	267280	14	AC114217 Rattus no
16	21.8	87.2	275074	14	AC111755 Rattus no
17	20.8	83.2	5119	9	AB011068 Rattus ra
18	20.8	83.2	75316	14	AP008202 Lotus cor

19	20.8	83.2	79081	14	AP008091	AP008091 Lotus cor
20	20.8	83.2	105384	15	AP006430	AP006430 Lotus cor
21	20.8	83.2	116810	8	AC112657	AC112657 Homo sapi
22	20.8	83.2	119118	8	HSR9760	AL049760 Human DNA
23	20.8	83.2	171747	9	AC132412	AC132412 Mus muscu
24	20.8	83.2	173818	5	BX323037	BX323037 Zebrafish
25	20.8	83.2	174224	14	AC148789	AC148789 Zebrafish
26	20.8	83.2	175991	14	AC118957	AC118957 Oryzotag
27	20.8	83.2	195768	8	AC022034	AC022034 Rattus no
28	20.8	83.2	195813	14	AC162431	AC162431 Homo sapi
29	20.8	83.2	225678	14	AC156884	AC156884 Papio ham
30	20.4	81.6	86584	14	AC166947	AC166947 Bos tauru
31	20.4	81.6	140741	8	HS298E2	AC169947 Bos tauru
32	20.4	81.6	142560	9	AC140436	AC163192 Homo sapi
33	20.4	81.6	156860	14	AC158834	AC140436 Mus muscu
34	20.4	81.6	169938	9	AC142406	AC158834 Bos tauru
35	20.4	81.6	177866	9	AC124352	AC142406 Mus muscu
36	20.4	81.6	180048	14	AC157436	AC124352 Mus muscu
37	20.4	81.6	204401	14	AC105524	AC157436 Oryzotag
38	20.4	81.6	205446	14	AC151152	AC105524 Rattus no
39	20.4	81.6	210179	14	AC128955	AC151152 Bos tauru
40	20.4	81.6	216422	14	AC096528	AC128955 Rattus no
41	20.4	81.6	220669	14	AC123017	AC096528 Rattus no
42	20.4	81.6	221791	9	AC162948	AC123017 Rattus no
43	20.4	81.6	237549	14	AC096159	AC162948 Mus muscu
44	20.4	81.6	239661	14	AC132157	AC096159 Rattus no
45	20.4	81.6	245581	14	AC111918	AC132157 Rattus no
46	20.4	81.6	255682	14	AC164434	AC111918 Rattus no
47	20.2	80.8	432	10	BV442141	AC164434 Mus muscu
48	20.2	80.8	492	10	BV442141	BV175580
49	20.2	80.8	579	10	BV340931	BV442141 S237P6151
50	20.2	80.8	582	10	BV159860	BV340931 S237P6151
51	20.2	80.8	639	10	BV374610	BV159860 RPDMMSE00
52	20.2	80.8	777	10	BV624515	BV374610 S231P6614
53	20.2	80.8	3638	5	BC084742	BV624515 S215P6064
54	20.2	80.8	13558	6	AX287041	BC084742 Xenopus 1
55	20.2	80.8	17893	6	AX346266	AX287041 Sequence
56	20.2	80.8	51858	8	AF097649	AX346266 Sequence
57	20.2	80.8	56646	8	AC117487	AF097649 Homo sapi
58	20.2	80.8	61798	8	AL512648	AC117487 Homo sapi
59	20.2	80.8	63637	14	AC135077	AL512648 Human DNA
60	20.2	80.8	66438	5	CR626939	AC135077 Homo sapi
61	20.2	80.8	66834	14	AC141798	CR626939 Zebrafish
62	20.2	80.8	67374	14	AC135336	AC141798 Apis mell
63	20.2	80.8	70282	14	AC111072	AC135336 Homo sapi
64	20.2	80.8	77521	14	AC018562	AC111072 Mus muscu
65	20.2	80.8	96268	8	AL607150	AC018562 Homo sapi
66	20.2	80.8	100887	15	ATAC011620	AL607150 Human DNA
67	20.2	80.8	108657	8	AL354921	ATAC011620 Arabidops
68	20.2	80.8	108759	1	CP000084_12	AL354921 Human DNA
69	20.2	80.8	110000	15	AP008214_061	Continuation (13 o
70	20.2	80.8	110000	15	AP008214_061	Continuation (5 of
71	20.2	80.8	113964	15	AC126010	Continuation (62 o
72	20.2	80.8	114084	5	BX942820	AC126010 Medicago
73	20.2	80.8	116326	5	BX323467	BX942820 Zebrafish
74	20.2	80.8	117636	8	HSJ365012	BX323467 Zebrafish
75	20.2	80.8	117818	14	AC160013	HSJ365012 Human DNA
76	20.2	80.8	122717	14	AP001186	AC160013 Medicago
77	20.2	80.8	123400	5	CR759881	AP001186 Homo sapi
78	20.2	80.8	125051	8	AL450164	CR759881 Zebrafish
79	20.2	80.8	125633	14	AC152867	AL450164 Human DNA
80	20.2	80.8	126150	8	AC005152	AC152867 Dasytus n
81	20.2	80.8	127688	15	AC137832	AC005152 Homo sapi
82	20.2	80.8	127824	8	AL592205	AC137832 Medicago
83	20.2	80.8	130977	8	AP002347	AL592205 Human DNA
84	20.2	80.8	134001	14	AC163754	AP002347 Homo sapi
85	20.2	80.8	136501	14	CR854932	AC163754 Loxodonta
86	20.2	80.8	137726	8	AC105919	CR854932 Danio rer
87	20.2	80.8	138878	14	AC160596	AC105919 Homo sapi
88	20.2	80.8	140376	14	AC162759	AC160596 Loxodonta
89	20.2	80.8	144249	14	AC011262	AC162759 Loxodonta
90	20.2	80.8	145318	8	AC084212	AC011262 Homo sapi
91	20.2	80.8	146759	9	AL604063	AC084212 Homo sapi

Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 27.1.

NOTE: This insert is not the entire sequence of the clone (entire sequence is 171.5kb). It is clipped at the overlaps with AC022493 and AC008434. The number of bases overlapped with AC022493 is 11961 and with AC008434 is 6448.

FEATURES

source Location/Qualifiers

1.88814
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-491N17"

ORIGIN

Query Match 93.6%; Score 23.4; DB 8; Length 88814;
Best Local Similarity 96.0%; Pred. No. 59;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
85877 AAAAAAAAAAGTCCCAATTCAGATA 85853

Db 85877 AAAAAAAAAAGTCCCAATTCAGATA 85853

RESULT 3
AC008434 103701 bp DNA linear PRI 01-OCT-2002
LOCUS Homo sapiens chromosome 5 clone CTC-325J23, complete sequence.
DEFINITION AC008434
AC008434.5 GI:23396211
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 103701)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 103701)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 103701)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
AUTHORS Submitted (01-OCT-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Oct 1, 2002 this sequence version replaced gi:13699352.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov

COMMENT

Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.4.
NOTE: Shatter libraries failed to resolve the dinucleotide repeat
region from 60278 to 61126. Forced join at 60885.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 136kb). It is clipped at the overlap with AC018764. The
number of bases overlapped is 17570.
Location/Qualifiers
1.103701
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-325J23"
60278..61126
/note="NOTE: Shatter libraries failed to resolve the

FEATURES

source Location/Qualifiers

1.103701
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-325J23"
60278..61126
/note="NOTE: Shatter libraries failed to resolve the

ORIGIN

Query Match 93.6%; Score 23.4; DB 8; Length 103701;
Best Local Similarity 96.0%; Pred. No. 57;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

dinucleotide repeat region from 60278 to 61126. Forced
join at 60885."

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
3523 AAAAAAAAAAGTCCCAATTCAGATA 3499

Db 3523 AAAAAAAAAAGTCCCAATTCAGATA 3499

RESULT 4
AC104467 163303 bp DNA linear HTG 26-FEB-2002
LOCUS Homo sapiens chromosome 3q clone RP11-21E14, WORKING DRAFT
DEFINITION AC104467
SEQUENCE 4 unordered pieces.
AC104467.3 GI:18874187
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE

1 (bases 1 to 163303)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.U.,
Fairhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harrie,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homes,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudas,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,C., Liu,W., Louisaed,H.,
Lozado,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nockenkwu,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Puyon,L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherrer,S., Scott,G., Shen,H., Shooshbari,N., Sleson,I.,
Sodergren,B., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telitod,B., Thomas,N., Thomas,S.,
Uemani,K., Vaequez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
TITLE 2 (bases 1 to 163303)
JOURNAL Worley,K.C.
REFERENCE Submitted (12-DEC-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
(bases 1 to 163303)

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (26-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 24, 2002 this sequence version replaced gi:17933806.

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

----- Project Information

Center project name: HCV

Center clone name: RP11-21E14

----- Summary Statistics

Sequencing vector: Plasmid: M77789

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 162741 bases at least Q40

Consensus quality: 163445 bases at least Q30

Consensus quality: 163933 bases at least Q20

Estimated insert size: 162788; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 7.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 102608: contig of 102608 bp in length
* 102609 102708: gap of unknown length
* 102709 142726: contig of 40018 bp in length
* 142727 142826: gap of unknown length
* 142827 160594: contig of 17766 bp in length
* 160595 160694: gap of unknown length
* 160695 163303: contig of 2609 bp in length.

FEATURES

Location/Qualifiers

1..163303

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="3q"

/clone="RP11-21E14"

102609..102708

/estimated_length=unknown

142727..142826

/estimated_length=unknown

160595..160694

/estimated_length=unknown

ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 163303;

Best Local Similarity 96.0%; Pred. No. 50;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 137304 AAAAAAAAAAGTCCCAATTCAGATA 137280

RESULT 5
LOCUS AY275681
DEFINITION Homo sapiens muts homolog 3 (E. coli) (MSH3) gene, complete cds.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AY275681
AY275681.1 GI:30089005

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 223885)

Rieder, M.J., Livingston, R.J., Daniels, M.R., Chung, M.-W.,

Miyamoto, K.E., Nguyen, C.P., Nguyen, D.A., Poel, C.L., Robertson, P.D.,

Schackwitz, W.S., Sherwood, J.K., Witrak, L.A. and Nickerson, D.A.,

Submitted (15-APR-2003) Genome Sciences, University of Washington,

1705 NE Pacific, Seattle, WA 98195, USA

To cite this work please use: NIEHS-SNPs, Environmental Genome

Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA

(URL: <http://egp.ge.washington.edu>).

FEATURES

Location/Qualifiers

source

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

REFERENCE 1 (bases 1 to 62860)
 AUTHORS Hominidae; Homo.
 TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
 JOURNAL Cambridgehire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

COMMENT

Direct Submission
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgehire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
 Clone requesters: clonerequest@sanger.ac.uk
 On Nov 16, 2001 this sequence version replaced gi:15131472.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMEP; Information
 on the WORMEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr9
 Rpl1-62L10 is from the library RPl1-11.1 constructed by the group
 of Piter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: PBACe3.6

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: vegas@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one subclone; and the assembly was confirmed by restriction digest,
 except on the rare occasion of the clone being a YAC.

FEATURES

SOURCE

Location/Qualifiers

1..62860
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RPl1-62L10"
 /clone_id="RPl1-11.1"

misc_feature
 1
 /note="Clone_left_end: RPl1-62L10"
 16471

misc_feature
 /note="Clone_right_end: RPl1-418K10"
 26383..26804

gene

/locus_tag="RPl1-62L10.1-001"

CDS

/locus_tag="RPl1-62L10.1-001"

/note="match: proteins: P38663"

/pseudo

/codon_start=1

60861
 /note="Clone_left_end: RPl1-80H12"

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 62860;
 Best Local Similarity 95.8%; Pred. No. 1.3e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24

Db 34383 AAAAAAAAAAGTCCCAATTCAGAT 34360

RESULT 7
 AC162929_1/c

WPCOMMENT

Sequence split into 4 fragments LOCUS AC162929 Accession AC162929
 Fragment Name Begin End

AC162929_0 1 110000
 AC162929_1 100001 210000
 AC162929_2 200001 310000
 AC162929_3 300001 392891
 Continuation (2 of 4) of AC162929 from base 100001 (AC162929 Mus musculus chromosome 9)

Query Match 89.6%; Score 22.4; DB 14; Length 110000;
 Best Local Similarity 95.8%; Pred. No. 1.3e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24

Db 58561 AAAAAAAAAAGTCCCAATTCAGAT 58538

RESULT 8

LOCUS

AC140409 200157 bp DNA linear ROD 27-JAN-2005
 DEFINITION Mus musculus BAC clone RP23-138K22 from 9, complete sequence.

ACCESSION

AC140409 GI:54292075

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

NOTICE:
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCT-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES
source
Location/Qualifiers
1..200157

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="9"
/map="9"
/clone="RP23-138K22"
/clone_1b="RPCT-23"
28950..28974
/note="Sequence derived from one plasmid subclone."
33610..33719
/note="Sequence derived from one plasmid subclone."
58180..58271
/note="Unresolved simple sequence repeat."
141326..141407
/note="Sequence derived from one plasmid subclone."
146736..146777
/note="Sequence derived from one plasmid subclone."
149164..149651
/note="Unresolved simple sequence repeat."
154697..164768
/note="Unresolved simple sequence repeat."
180921..180989
/note="Sequence derived from one plasmid subclone."
190161..190465
/note="Unresolved tandem repeat."

ORIGIN

Query Match 89.6%; Score 22.4; DB 9; Length 200157;
Best Local Similarity 95.8%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAAGT 24
|||||

DB 133974 AAAAAAAAAAGTCCCAATTCAAGT 133997

RESULT 9
AX348364/c
LOCUS AX348364 5933 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 59 from Patent WO0202806.
ACCESSION AX348364
VERSION AX348364.1 GI:18614400
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Method and nucleic acids for pharmacogenomic methylation analysis
JOURNAL Patent: WO 0202806-A 59 10-JAN-2002;
Epigenomics AG (DE)
LOCATION/Qualifiers
1..5933
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"

ORIGIN /note="Chemically treated genomic DNA (Homo sapiens)"

Query Match 87.2%; Score 21.8; DB 6; Length 5933;
Best Local Similarity 92.0%; Pred. No. 5e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAAGT 25
|||||

DB 2331 AAAAAAAAAAGTCCCAATTCAAGT 2307

RESULT 10
AX347335/c
LOCUS AX347335 10151 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2406 from Patent WO0200928.
ACCESSION AX347335
VERSION AX347335.1 GI:18495223
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.

REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2406 03-JAN-2002;
Epigenomics AG (DE)
LOCATION/Qualifiers
1..10151

FEATURES
source
Location/Qualifiers
1..10151
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 87.2%; Score 21.8; DB 6; Length 10151;
Best Local Similarity 92.0%; Pred. No. 4.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAAGT 25
|||||

DB 2331 AAAAAAAAAAGTCCCAATTCAAGT 2307

RESULT 11
AC007067/c
LOCUS AC007067 89934 bp DNA linear PLN 28-JUN-2000
DEFINITION Genomic sequence for Arabidopsis thaliana BAC T10024 from Chromosome 1, complete sequence.
ACCESSION AC007067
VERSION AC007067.4 GI:4558521
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 89934)
Shinn, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Khan, S., Kim, C., Walker, M., Altafi, H., Araujo, R., Conn, L., Conway, A. B., Gonzalez, A., Hansen, N. F., Hultzer, L., Kremetschke, I., Lenz, C., Li, J., Liu, S., Luros, S., Rowley, D., Schwartz, J., Tortum, M., Vysotskaya, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and Ecker, J. R.
Genomic sequence for Arabidopsis thaliana BAC T10024 from Chromosome 1
Unpublished
2 (bases 1 to 89934)
Ecker, J. R.
Direct Submission
Submitted (12-MAR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 89934)

REFERENCE AUTHORS	4 (bases 1 to 99416) Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C., Shinn, P., Altafi, H., Bei, B., Chin, C., Chou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharasy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vayberg, M., Yu, G., Davis, R., Federapfel, N., Theologis, A. and Becker, J.
TITLE	Submitted (14-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE AUTHORS	5 (bases 1 to 99416) Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C., Shinn, P., Altafi, H., Bei, B., Chin, C., Chou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharasy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vayberg, M., Yu, G., Davis, R., Federapfel, N., Theologis, A. and Becker, J.
TITLE	Direct Submission
JOURNAL	Submitted (14-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT	Submitted (11-OCT-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
FEATURES	On Nov 6, 1999 this sequence version replaced gi:6137668.
SOURCE	Location/Qualifiers 1..99416 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="1" /clone="F20B24" complement(join(2528..2776,2941..3237)) /note="similar to tyrosyl-tRNA synthetase isoolog gi 1707012; similar to BGTS gb 143306, gb A196935.1, gb H76501, gb AA605326, gb H65365" /codon_start=1 /evidence=not_experimental /product="F20B24.1" /protein_id="AAF17673.1" /db_xref="GI:6573753" /translation="MLLPANFCRVYIKAAEATGALRKVPFTKNELRPGTNGSLNVN KYVTRKYNQGRGGRPMQGRQROMRIAEGLVDGDTGIIIFTRANGQGEINLFWYFSS CTFYRPTWDALIHFTVDMMKESGVTLNNAKIDMKSGMRLAVDRMGVGEVAAEFTD ITVXDDNNLSLIEVLVSVEY" join(3969..4197,4232..4317,4372..4596,4630..4692, 4773..5040,5099..5394) /note="similar to AMSH gi 4098124" /codon_start=1 /evidence=not_experimental /product="F20B24.2" /protein_id="AAF17652.1" /db_xref="GI:6573732" /translation="WVTLSPPSLSCVENVTCKSHVSRVLISGTDINHGSESAK ILRDHIVITLTFPESQPSVYVYQARSNLKRLLDTELRARENTEDELFCGCLA ATLNLGGEKIKSNVRKFTLLITLITSNRDLCKFEVLKLDLCYCKLTFPCPGFCHD SWLFFGCSKIVLSSFMNIMERGIFVYVTLIIKPESTNSQCANNEVAVSLQRE LVPVMHIVYLLBLAPMSFSDLFIVYLVYEMWHIIPOTHSQCFMSVLDLTHYS YOVHLCFOIAYILFAIVAPVTPSSKVLVLMNSQLSHRISLISQTLFERSYGV IKFLNDPGMEVLRGSETRGFPHKPEEDGNVYEHCSVYKNSNLRFEIFDLR" 6127..6402 /note="similar to C. elegans AL2 homolog gi 1465851" /codon_start=1 /evidence=not_experimental /product="F20B24.3" /protein_id="AAF17653.1" /db_xref="GI:6573733" /translation="WGFIRKKAASSPHIVCCYQNPYRSTLPCFSLXGRFOAHDSF LRTVSKTSLSLSTVTVTDLGTETRRRFTLLSFLSGF" join(6667..6901,7022..7194) /note="similar to PDR5-like ABC transporter emb CAA94437" /codon_start=1 /evidence=not_experimental

/product="F20B24.4"
/protein_id="AAFI7654.1"
/db_xref="GI:6573734"
/translation="MMMRGGERVEKRLRPFVDSRTWDLCVIWKLGDDPSRFNTLSGPF
SLFPRCSCESSAVYFVSVAKLPFLCILMGSPFLNFMWCSIKFLMPLFFPFAHL
VFPNGMLFVLYHLPPVGLVLSFPTKSLMVL"
join(7417, 7551, 7803, 8282, 8305, 8808)
/note="similar to phaseolin G-box binding protein PGI
gi|1142619"
/codon_start=1
/evidence=not experimental
/product="F20B24.5"
/protein_id="AAFI7655.1"
/db_xref="GI:6573735"
/translation="MPYIPGVISYISPPFQVNFPPFKNMKLLSQLQDSWRSSDVKISK
VRPDESVMHLIMSKCTTFEPPEQRLQFRIIPRAESMSGVNLSVEGGSSSVN
PSESTONLFGNYPNASCEILREOTPLIMNKEDVYQVONANDSKANKLLPTENK
SKINHSRKRERINQAMGLRAVAPKITVSSETOCSPITLLVDYIYLSLCYML
NKIGIFSDAVYINELVEKQLEDEKINMECKELAAEEOAIAIDPRAERSSKS
NKRKKEVLEPYTIDLDNLKLSCKIPQTLVSFOYKLEHETGERDPLIRVDEHKOD
GFKRLFAVDCELEITIDVNFRLDLTVMTVLANKVY"
complement(join(9205, 9447, 9545, 9706, 9803, 9950,
10233, 10308, 10435, 10521, 10620, 11123, 11252, 12179))
/note="Putative serine/threonine protein kinase
emb|CAA18823.1"
/codon_start=1
/evidence=not experimental
/product="F20B24.6"
/protein_id="AAFI7672.1"
/db_xref="GI:6573752"
/translation="MDKVCQOQADLFEGTISPFVASOPTNVGFTDQKITGSETTOP
ATSPSPSPDPTQSPPPATADPPNPQPNPTTTPPPSPSPSPSPSPSPSPSPSP
PQSTPTDSFVVIPEPKQLPPBSLPFPPLVQQLPPRPNDNNITLPPSPSPSP
PSP
SP
IAGVFYRRKOKKSSGSSPRNOYLPPANVSVNTEGFHYHOKGKNSSQNSDPDN
SLGNPKHGRGTPDQSAVGTCTKHITVTELSQITHEGCKSFVYVGGCGCYKGLPBG
KPAVIAIKOKSVASAGREFAEVETISRHHRHLSLVGTCISQHPFLIYEPVNT
LDHLHGTIFFLQSLFVTKLGFYIEVCKNLPLYELWSRVRIAGAAGLAIYEDCH
PKIIRHDKSSNLLDDEFAQAKYLAPEVASSGKLDRSDRVRVIAIGAAGLAIYEDCH
VDTGSPGSESLVEMAPRLIEAIKEDISEVVDPRLENDYSEVYKMEYVASCVR
HSAKRPWQVVALDTRDLDLDTNGVAGQSRVYDSQYSENEIRIFPARABSDSD
LGNTGYPSQDVAATSHVESERATSHRN"
complement(join(13534, 13636, 13729, 13847, 13934, 14053,
14139, 14209, 14287, 14434))
/note="similar to ADP-ribosylation factor gb|AAD17207;
similar to ESTs gb|N65779, gb|A1996946.1, gb|A1725909.1,
gb|T45928, gb|T42202"
/codon_start=1
/evidence=not experimental
/product="F20B24.7"
/protein_id="AAFI7671.1"
/db_xref="GI:6573751"
/translation="MGISPAKLFSPRLFAKEMRLIWLGLDAAGTIIYKLGELVT
TITIGPNVETVEYKNTSFTVWVGQDCKRPIRMRYFQTOGLIFVDSNDRIIVE
ARELIRMLNEDLRDAVLVFAKQKDLPPANMAAETDQLGHSIRQRMWYIQSTCA
TSEGLYEGLDWISNNIAKVRLLNS"
complement(join(15682, 16265, 16344, 16551, 16591, 16665,
16754, 16975, 17052, 17183, 17604, 18011))
/note="similar to polygalacturonase-like protein
emb|CAA66811; similar to ESTs gb|R29820, gb|A1993509.1,
emb|F19830"
/codon_start=1
/evidence=not experimental
/product="F20B24.8"
/protein_id="AAFI7670.1"
/db_xref="GI:6573750"
/translation="MIRKSLRSITTMILAAVLVWSVTLCTIARGRHWHSHRS
SSDLSLSLKKRSH
PKVQPPPLPPQKSGOVFNWMDFGAKGCDCTKAEAAVMAKLEASMLVPPVYT
YLVGPIFSFGPQCANIVFOLDGTIIAPTDSKTGKGLMMWIDFTLKGIGKYQKGV
DGGSGMWQODSPFIDSDTKLIVPLNNSANONPMPIRSLDRMSISIPYANRW
CEIALAFSGSFGVEVIGITIQNSPQCHKLFKDDCVGVVADIANVSGDSBPNDTGHQ

CDS
/product="F20B24.9"
/protein_id="AAFI7669.1"
/db_xref="GI:6573749"
/translation="MLSGNNGTAPPVFNENQLOQTNRSLNQLHLGMMGGCTVD
PNVYFANDNLVPMIRPKRGRASISNNIQRQOQLOMSLANNYNNTSVREVEPKEN
LVSTGRLSYDDDEHNSVTSAGSIIAASPIFOSLDDSLRIHLHQKDFDQFIKO
VLVSCRLCFEYKRPFDSPVFCFVVOAQAOKAKRYKDMKQKHITSLTLKGVSK
KLOEKHEINDMKRKKELVERIKQVAMEQNMHYAKTINSEYVNTLKAALQMSHN
NSTIAADQSGEJGSEIDDAASSTIIDNNNNNNMGIHQRRCCKMCKVSESVLIV
PCRLSLCKECDFETKICPCKSLKSKSVQVPS"
/codon_start=1
/evidence=not experimental
/product="F20B24.9"
/protein_id="AAFI7669.1"
/db_xref="GI:6573749"
/translation="MLSGNNGTAPPVFNENQLOQTNRSLNQLHLGMMGGCTVD
PNVYFANDNLVPMIRPKRGRASISNNIQRQOQLOMSLANNYNNTSVREVEPKEN
LVSTGRLSYDDDEHNSVTSAGSIIAASPIFOSLDDSLRIHLHQKDFDQFIKO
VLVSCRLCFEYKRPFDSPVFCFVVOAQAOKAKRYKDMKQKHITSLTLKGVSK
KLOEKHEINDMKRKKELVERIKQVAMEQNMHYAKTINSEYVNTLKAALQMSHN
NSTIAADQSGEJGSEIDDAASSTIIDNNNNNNMGIHQRRCCKMCKVSESVLIV
PCRLSLCKECDFETKICPCKSLKSKSVQVPS"
Query Match 87.2%; Score 21.8; DB 15; Length 99416;
Best Local Similarity 92.0%; Pred. No. 2,1e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 6945 AAAAAAAAAAGTCTAATTCAGAA 6921
RESULT 13
AC160146/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
NTKDVLIHSTTACGGDCJISIQGCSNVFPHANNCGPHGISITGSLCKEGTKASNI
TVRDVAHNTMTGVRIRKTWQGVGVGIIIESNIQLNQVQIPITINPEYCDHSCXKQ
TSAVAEGVYIERIKGYIVKPHFACSDNPPCVQDLSIELKPVQEKYRMAYCWC
QTEGELNTPPIIDCKIKGPPRNKQSDHYC"
complement(join(24326, 25323, 25414, 25522))
/note="similar to putative inhibitor of apoptosis
gi|3924605; similar to ESTs gb|A1994578.1, gb|T04171,
gb|A1728525"
/codon_start=1
/evidence=not experimental
/product="F20B24.9"
/protein_id="AAFI7669.1"
/db_xref="GI:6573749"
/translation="MLSGNNGTAPPVFNENQLOQTNRSLNQLHLGMMGGCTVD
PNVYFANDNLVPMIRPKRGRASISNNIQRQOQLOMSLANNYNNTSVREVEPKEN
LVSTGRLSYDDDEHNSVTSAGSIIAASPIFOSLDDSLRIHLHQKDFDQFIKO
VLVSCRLCFEYKRPFDSPVFCFVVOAQAOKAKRYKDMKQKHITSLTLKGVSK
KLOEKHEINDMKRKKELVERIKQVAMEQNMHYAKTINSEYVNTLKAALQMSHN
NSTIAADQSGEJGSEIDDAASSTIIDNNNNNNMGIHQRRCCKMCKVSESVLIV
PCRLSLCKECDFETKICPCKSLKSKSVQVPS"
AC160146 167510 bp DNA linear ROD 02-JUN-2005
(Mus musculus 10 BAC R224-421112 (Roswell Park Cancer Institute
C57BL/6J Male) Mouse BAC library) complete sequence.
AC160146 AC120407
AC160146.10 GI:68533269
HTG.
Mus musculus (house mouse)
Mus musculus
Bukayova; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 167510)
Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooke,S., Archer,P.,
Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
Berducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chantour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denison,S., Devamo,C., Ding,Y., Dinh,H.,
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dzinda,E., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaskin,C., Gench,S., Ghose,S., Giller,R., Gonzalez,D.,
Gonzalez-Garay,M., Guevara,W., Holder,M., Haland,W., Haebleren,K.,
Hall,B., Hamid,H., Hamilton,K., Harber,B., Harris,R., Havlak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgson,A., Hognes,M., Hollins,B.,
Howell,L.T., Huliy,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafas,K., Kelly,S.,
Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,
Leal,S., Lee,K., Lee,S., Legall,F.I., Lemon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linell,M., Liu,W., Liu,Y.-S., Liu,Y., Llyanage,D.,
London,P., Lopez,J., Lorensheva,L., Lozdo,R., Luk,T., Madu,R.,
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
McClelland,H., McPherson,J., Mercedao,C., Metzker,M.,
Mulasavijevic,A., Minja,E., Morgan,M., Morris,S., Mundasa,M.,
Murray,D., Nazareth,L., Ngo,D., Nguyen,N., Norwig-Fasbaugh,E.,
Nott,A., Naocheleleh,O., Obregon,M., Och-Okorle,C., Odeh,E.,
Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,

Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Phan,T.,
Primus,E., Pu,L.-L., Puazo,M., Qin,X., Qiu,M., Quito,J.,
Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sama,M., Sanders,W.,
Sanibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
Stason,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,
Triejos,Z., Usmani,K., Vargo,C., Verdusco,D., Villaana,D., Virk,D.,
Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,
Zhang,Z., Zhou,J., Weinstein,G. and Gibbs,R.

JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 167510)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 167510)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 167510)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2005) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jul 2, 2005 this sequence version replaced g1:68158960. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: WIBR
Contact: hgsc-help@bcm.tmc.edu.
Location/Qualifiers
1. 167510
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="10"
/clone="RP24-421112"
1. 61466
/note="overlaps bases 127991..189457 of clone AC135019"
/function="clone overlap"
repeat_region 169..236
/rpt_family="CT-rich"
repeat_region complement(238..297)
/rpt_family="B1_Mur2"
repeat_region 330..437
/rpt_family="MERSA"
repeat_region 708..786
/rpt_family="PB1D7"
repeat_region 973..1109
/rpt_family="B1_Mur4"
repeat_region complement(3682..3822)
/rpt_family="B1F2"

repeat_region complement(4188..4326)
/rpt_family="B1_Mur3"
repeat_region complement(4398..4499)
/rpt_family="PB1D10"
repeat_region 4677..4720
/rpt_family="(CA)n"
4679..4720
/note="variable size of repeat in subclones, size is 22-21AC"
misc_feature function="variable size of repeat between subclones"
4958..5061
/rpt_family="PB1D10"
repeat_region 5080..5228
/rpt_family="ID-B1"
repeat_region complement(7578..7674)
/rpt_family="PB1D7"
8111..8149
/rpt_family="(CGTG)n"
10000..10257
/rpt_family="B4"
10314..10388
/rpt_family="PB1D10"
repeat_region complement(10612..10678)
/rpt_family="PB1D10"
complement(10956..11075)
/rpt_family="B3"
complement(12182..12300)
/rpt_family="B1_Mus1"
12653..12767
/rpt_family="PB1D9"
complement(12961..13019)
/rpt_family="ID4"
13029..13173
/rpt_family="MERS8A"
13295..13434
/rpt_family="B1_Mus1"
13514..13556
/rpt_family="(CA)n"
13566..13681
/rpt_family="GA-rich"
14902..15104
/rpt_family="B3"
complement(15458..15692)
/rpt_family="B3"
complement(15784..15844)
/rpt_family="RSINE1"
16255..16507
/rpt_family="B4A"
16829..16931
/rpt_family="B1_Mus2"
complement(17045..17146)
/rpt_family="PB1D9"
17195..17384
/rpt_family="B3A"
17394..17496
/rpt_family="B1F1"
17555..17672
/rpt_family="MT2B2"
17582..17682
/rpt_family="PB1D10"
17683..17758
/rpt_family="(CAA)n"
complement(18240..18376)
/rpt_family="B3A"
complement(18420..18572)
/rpt_family="RSINE1"
19018..19038
/rpt_family="AT-rich"
19230..19316
/rpt_family="PB1"
19696..19897
/rpt_family="B3"
19910..20095

```

repeat_region      /rpt_family="B3"
                    complement(20261..20347)
repeat_region      /rpt_family="charlieta"
                    20524..20605
repeat_region      /rpt_family="PBD10"
                    complement(21637..21753)
repeat_region      /rpt_family="ID_B1"
                    complement(22601..22804)
repeat_region      /rpt_family="B3A"
                    23063..23085
repeat_region      /rpt_family="GC_rich"
                    complement(23200..23277)
repeat_region      /rpt_family="ID4_"

```

```

Query Match      87.2%  Score 21.8; DB 9; Length 167510;
Best Local Similarity 92.0%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCATTCAGATA 25
Db 39619 AGAAAAAAAGTCACATTCAGATA 39595

```

```

RESULT 14
AC135019/c 189457 bp DNA linear ROD 27-JAN-2005
LOCUS Mus musculus BAC clone RP23-186D15 from 10, complete sequence.
AC135019
VERSION AC135019.3 GI:51854778
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Kruchowski, S., Haglund, K., Bielicki, L. and Hakenson, W.
TITLE The sequence of Mus musculus BAC clone RP23-186D15
JOURNAL Unpublished (2001)
AUTHORS 2 (bases 1 to 189457)
REFERENCE McPherson, J.D. and Waterston, R.H.
AUTHORS Direct Submision
TITLE Submitted (04-OCT-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA
AUTHORS 3 (bases 1 to 189457)
REFERENCE Wilson, R.K.
AUTHORS Direct Submision
TITLE Submitted (10-JUL-2004) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA
AUTHORS 4 (bases 1 to 189457)
REFERENCE Wilson, R.K.
AUTHORS Direct Submision
TITLE Submitted (02-SEP-2004) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA
AUTHORS 5 (bases 1 to 189457)
REFERENCE Wilson, R.K.
AUTHORS Direct Submision
TITLE Submitted (27-JAN-2005) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 2, 2004 this sequence version replaced gi:49458034.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_BA0186D15

```

NOTICE:

This sequence was finished as follows unless otherwise noted:

FEATURES

source

```

repeat_region      /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="10"
                    /map="10"
                    /clone="RP23-186D15"
                    /clone_11b="RP23-23"
                    319..461
repeat_region      /rpt_family="Alu"
                    549..676
repeat_region      /rpt_family="Alu"
                    677..824
repeat_region      /rpt_family="Alu"
                    825..835
repeat_region      /rpt_family="Alu"
                    991..1165
repeat_region      /rpt_family="B2"
                    2398..2471
repeat_region      /rpt_family="Alu"
                    2484..2635
repeat_region      /rpt_family="Alu"
                    2825..3012
repeat_region      /rpt_family="B2"
                    3360..3438
repeat_region      /rpt_family="ID"
                    3469..3616
repeat_region      /rpt_family="B2"
                    4104..4146
repeat_region      /rpt_family="B2"
                    4184..4612
repeat_region      /rpt_family="Y-chromosomal"
                    4279..4312
repeat_region      /rpt_family="MER1_type"
                    4313..4461
repeat_region      /rpt_family="Alu"
                    4462..4626
repeat_region      /rpt_family="MER1_type"
                    4883..5054
repeat_region      /rpt_family="B2"
                    5794..5946
repeat_region      /rpt_family="Alu"
                    6322..6671
repeat_region      /rpt_family="MALR"
                    7644..7757
repeat_region      /rpt_family="Alu"
                    7787..7926
repeat_region      /rpt_family="Alu"
                    8167..8262

```

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

```

repeat_region      /rpt_family="ERVX"
                    8428..8586
repeat_region      /rpt_family="B2"
                    8615..8713
repeat_region      /rpt_family="Alu"
                    8737..8923
repeat_region      /rpt_family="B2"
                    8975..9049
repeat_region      /rpt_family="ID"
                    9069..9149
repeat_region      /rpt_family="Alu"
                    9153..9352
repeat_region      /rpt_family="B2"
                    9538..9932
repeat_region      /rpt_family="MALR"
                    9960..10121
repeat_region      /rpt_family="B4"
                    10177..10243
repeat_region      /rpt_family="B4"
                    10458..10567
repeat_region      /rpt_family="B4"
                    10728..10880
repeat_region      /rpt_family="B2"
                    11397..11569
repeat_region      /rpt_family="B2"
                    12144..12412
repeat_region      /rpt_family="MER2_type"
                    13013..13160
repeat_region      /rpt_family="Alu"
                    13229..13375
repeat_region      /rpt_family="Alu"
                    13384..13488
repeat_region      /rpt_family="Alu"
                    13490..13675
repeat_region      /rpt_family="B2"
                    14257..14427
repeat_region      /rpt_family="B2"
                    15050..15263
repeat_region      /rpt_family="B2"
                    15539..15721
repeat_region      /rpt_family="B2"
                    16086..16303
repeat_region      /rpt_family="B2"
                    16378..16597
repeat_region      /rpt_family="L1"
                    16604..16783
repeat_region      /rpt_family="B2"
                    complement(16707..16779)
tRNA
                    /note="Likely pseudogene (HMM Sc=30.10 / Sec struct
                    Sc=-7.10)"
repeat_region      16875..17021
                    /rpt_family="Alu"
repeat_region      17037..17126
                    /rpt_family="U6"
misc_feature        18602..19659
                    /note="Cpg island (GC=70.4, o/e=1.04, #CpGs=125)"
repeat_region      19912..20071
                    /rpt_family="B2"
repeat_region      20803..20923
                    /rpt_family="Alu"
repeat_region      21062..21193
                    /rpt_family="Alu"
repeat_region      21275..21422
                    /rpt_family="B4"
repeat_region      21423..21570
                    /rpt_family="Alu"
repeat_region      21571..21651
                    /rpt_family="B4"
repeat_region      21704..21801
                    /rpt_family="Alu"
repeat_region      21725..21816
                    /rpt_family="B4"

```

```

repeat_region      21902..22023
                    /rpt_family="Alu"
repeat_region      22618..22804
                    /rpt_family="B2"
repeat_region      23045..23232
                    /rpt_family="MALR"

Query Match      87.2%; Score 21.8; DB 9; Length 189457;
Best Local Similarity 92.0%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTCCCAATTCAGATA 25
DB      167610 AGAAAAAAGTCACATTCAGATA 167586

RESULT 15
AC114217/c
LOCUS      267280 bp DNA linear HTG 10-OCT-2002
DEFINITION Rattus norvegicus clone CH230-316A18, *** SEQUENCING IN PROGRESS
ACCESSION AC114217
VERSION AC114217.3 GI:23101120
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridae; Muridae; Murinae; Rattus.
          1 (bases 1 to 267280)

REFERENCE
AUTHORS Muzny,D,Marie, Metzker,M,lee, Abramson,S, Adams,C, Alder,J,
          Allen,C, Allen,H, Alsdorke,S, Amth,A, Anguiano,D,
          Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
          Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benhmed,F,
          Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
          Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
          Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,
          Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
          Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
          Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedrich,D,
          Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinh,H, Dlyva,K,
          Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K,
          Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
          Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
          Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
          Gebregeorgis,E, Geer,K, Gill,R, Gladys,M, Guerra,W, Guevara,W,
          Guneratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
          Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
          Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogue,M,
          Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A,
          Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
          Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
          Kowls,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
          Liu,J, Liu,M, Liu,Y, London,P, Longacre,S, Lopez,J,
          Lorensunewa,L, Louised,H, Lozano,R,J, Lu,X, Ma,J,
          Maheshwari,M, Mahindaratne,M, Mamoud,M, Mallory,K, Mangum,A,
          Mangum,B, Mapua,P, Martin,K, Martin,R, Matley,E,
          Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
          Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
          Morgan,M, Morris,K, Morris,S, Munday,M, Murphy,M, Nair,L,
          Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
          Nwokediemen,O, Okunodu,G, Olarnpunsagoon,A, Pal,S, Parks,K,
          Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankuch,C,
          Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L,-L,
          Puazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
          Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
          Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,U,
          Sanders,W, Savery,G, Scherer,S, Scott,G, Shutsman,S, Shen,H,
          Shetty,J, Shvartbeyn,A, Sisson,I, Sitter,C,D, Smajs,D,
          Sneed,A, Sodergren,E, Song,X,-Z, Sorelle,R, Sosa,J,
          Steidle,M, Strong,R, Sutton,A, Svatek,A, Tabor,P, Taylor,C,
          Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umanji,K,
          Valae,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,J,

```

Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Weizsäcker, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Meisner, G. and Gibbs, R. A.

TITLE
Direct Submission

REFERENCE
2 (bases 1 to 267280)

AUTHORS
Worley, K. C.

JOURNAL
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
3 (bases 1 to 267280)

AUTHORS
Rat Genome Sequencing Consortium.

TITLE
Direct Submission

JOURNAL
Submitted (10-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Sep 18, 2002 this sequence version replaced gi:21738326. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GOME
Center clone name: CH230-316A18

Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 231500 bases at least Q40
Consensus quality: 234123 bases at least Q30
Consensus quality: 236133 bases at least Q20
Estimated insert size: 239449; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1..267280
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-316A18"
1..1241
/note="wgs contig"
156303..160898

misc_feature
156303..160898

misc_feature
/note="wgs contig"
complement(237875..238733)
/note="clone_boundary"
clone_end:T7
site:RhoI

misc_feature
end_sequence:RXMAN09TU"
23677..24139
/note="wgs_end_extension"
clone_end:T7"

misc_feature
/note="wgs_end_extension"
245676..247694
clone_end:T7"

gap
247695..247794
/estimated_length=unknown
247795..249658
/note="wgs_end_extension"
clone_end:T7"

misc_feature
249705..254184
/note="wgs_end_extension"
clone_end:T7"

gap
254185..254284
/estimated_length=unknown

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 267280;
Best Local Similarity 92.0%; Pred. No. 1.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 24473 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
AAAAAAAAAAAAAAAAAGTCCCAATTAAGATA 24449

RESULT 16
AC111755/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-248N13, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC111755.4 GI:24941988
VERSION
AC111755.4 GI:24941988
ACCESSION
AC111755.4 GI:24941988
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Buthyocca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 275074)

REFERENCE
AUTHORS
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alspbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhmed, F., Bismalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Cavazos, I., Caesar, H., Center, A., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Healand, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hu, Y. K., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshew, L., Louised, H., Lozada, R. D., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G., Olamunagbon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pflanz, C., Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L., L., Puzos, M., Quito, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, U., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Soes, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willison, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE
Journal
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 275074)
Worley, K.C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 275074)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23602290.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Project name: GOAP
Center project name: CH230-248N13
Center clone name: CH230-248N13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 250297 bases at least Q40
Consensus quality: 254244 bases at least Q30
Consensus quality: 256359 bases at least Q20
Estimated insert size: 253751; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1. 275074
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10115"
/clone="CH230-248N13"
1. 1503
/note="Wgs end_extension
clone end: T7"
103487. 103826
/note="clone_boundary
clone_end: T7"
site:
end_sequence: B2193680"
complement(271724..272582)
/note="clone boundary
clone_end: Sp6
site:
end_sequence: B2193682"
273811. 273910
/estimated_length=unknown

ORIGIN
gap

Query Match 87.2%; Score 21.8; DB 14; Length 275074;
Best Local Similarity 92.0%; Pred. No. 1.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY
1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
Db 233409 AAAAAAAAAAGTCCCAATTCAGATA 233385

RESULT 17
AB011068/c
LOCUS
DEFINITION
Rattus rattus mRNA for type II iodothyronine deiodinase, complete cds.
ACCESSION
AB011068
VERSION
AB011068.1 GI:2959448
KEYWORDS
type II iodothyronine deiodinase.
SOURCE
Rattus rattus (black rat)
ORGANISM
Rattus rattus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (sites)
Gondou, A., Toyoda, N., Nishikawa, M., Tabata, S., Yonemoto, T., Ogawa, Y., Tokoro, T., Sakaguchi, N., Wang, F. and Inada, M.,
Induction of type 2 deiodinase activity by cyclic guanosine 3',5'-monophosphate in cultured rat glial cells
Thyroid 8 (7), 615-622 (1998)
9709916
2 (bases 1 to 5119)
Toyoda, N.
Direct Submission
Submitted (12-FEB-1998) Nagasaki Toyoda, Kansei Medical University, 2nd Dept. of Internal Medicine; 10-15, Fumizono-cho, Moriyuchi, Osaka 560, Japan (E-mail: toyoda@takii.kmu.ac.jp, Tel: 81-6-992-1001, Fax: 81-6-996-4870)

REFERENCE
PUBMED
2
2 (bases 1 to 5119)
Toyoda, N.
Direct Submission
Submitted (12-FEB-1998) Nagasaki Toyoda, Kansei Medical University, 2nd Dept. of Internal Medicine; 10-15, Fumizono-cho, Moriyuchi, Osaka 560, Japan (E-mail: toyoda@takii.kmu.ac.jp, Tel: 81-6-992-1001, Fax: 81-6-996-4870)

TITLE
JOURNAL

FEATURES
source
1. 5119
/organism="Rattus rattus"
/mol_type="mRNA"
/db_xref="taxon:10117"
1. 801
/note="Ega codon at nucleotide 787. 789 may code for

selenocysteine or termination, referred to U53505"
/codon_start=1
/transl_except=(pos:388..390,aa:OTHER)
/transl_except=(pos:787..789,aa:OTHER)
/product="Type II iodothyronine deiodinase"
/protein_id="BAA25186.1"
/db_xref="GI:2970048"
/translation="MWGLSVLLTLTQTLVPFSPNCLFLAYDSVILLKRVALLSS
KSTGEMRMATSGELRCVWNSFLIDAKYKLEDDPNSVSVHNSPEAGNCSK
TADGAECILDFACAEPLVVFSGATXPPPTQLPAFRLVVEFSSVADFLVYIDE
AHPSDGAIVPDSSWSEFEKGRNQEDCAAHQLLEFSLPQCCQVADRMNNNAV
AYGVAFERVICVQRKLIAYLGKGPFSYNLGEVSMLEKNSKXIIID"

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 5119;
Best Local Similarity 91.7%; Pred. No. 1.2e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 2042 AAAAAAAAAAGTCCCAATTCAAAT 2019

RESULT 18
AP008202/c
LOCUS AP008202 75316 bp DNA linear HTG 28-DEC-2004
DEFINITION Lotus corniculatus var. japonicus chromosome 5 clone LjT09M14, ***
SEQUENCING IN PROGRESS ***, 5 unordered pieces.
ACCESSION AP008202
VERSION AP008202.1 GI:56806508
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.

REFERENCE 1
AUTHORS Kaneko, T., Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Structural Analysis of a Lotus japonicus Genome. XI. Sequence
JOURNAL Features and Mapping of Nine hundred twenty-one TAC Clones
REFERENCE 2 (bases 1 to 75316)
AUTHORS Sato, S.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research: 2-6-7 Kazusa-kamatari, Kisarazu,
Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
URL: http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
Fax:81-438-52-3934)

COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Source
1 .75316
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"

/chromosome="5"
/clone="LjT09M14"
/clone_lib="LjT library"
/note="TAC clone: TW1662, synonym: Lotus japonicus"
3841..3940
gap /estimated_length=unknown
5572..5671
gap /estimated_length=unknown
18680..18779
gap /estimated_length=unknown
59148..59247
gap /estimated_length=unknown

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 75316;
Best Local Similarity 91.7%; Pred. No. 5.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 38330 AAAAAAAAAAGTCCCAATTCAGATA 38307

RESULT 19
AP008091
LOCUS AP008091 79081 bp DNA linear HTG 28-DEC-2004
DEFINITION Lotus corniculatus var. japonicus chromosome 5 clone LjB26H09, ***
SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION AP008091
VERSION AP008091.1 GI:56806398
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.

REFERENCE 1
AUTHORS Kaneko, T., Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE Structural Analysis of a Lotus japonicus Genome. XI. Sequence
JOURNAL Features and Mapping of Nine hundred twenty-one TAC Clones
REFERENCE 2 (bases 1 to 79081)
AUTHORS Sato, S.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research: 2-6-7 Kazusa-kamatari, Kisarazu,
Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
URL: http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
Fax:81-438-52-3934)

COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Source
1 .75316
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"

```

FEATURES
*
65339 65438: gap of unknown length
65439 79081: config of 13643 bp in length.
Location/Qualifiers
1. 79081
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"
/chromosome="5"
/clone_lib="LjB26H09"
/clone_1ib="LjB 1library"
/note="BAC clone: BM1452, synonym: Lotus japonicus"
9454. 9553
gap /estimated_length=unknown
10386. 10485
gap /estimated_length=unknown
11424. 11523
gap /estimated_length=unknown
13464. 13563
gap /estimated_length=unknown
24033. 24132
gap /estimated_length=unknown
35157. 35256
gap /estimated_length=unknown
47506. 47605
gap /estimated_length=unknown
65339. 65438
gap /estimated_length=unknown

ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 79081;
Best Local Similarity 91.7%; Pred. No. 5.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2
AAAAAAAAAGTCCCAATTCAGATA 25
|||||
32248 AAAAAAAAAATCCCAATTCAGATA 32271

RESULT 20
AP006430 105384 bp DNA linear PLN 22-JUL-2003
LOCUS AP006430
DEFINITION Lotus corniculatus var. japonicus genomic DNA, chromosome 5,
clone: LjT33L13, TM0327, complete sequence.
ACCESSION AP006430
VERSION AP006430.1 GI:31581061
KEYWORDS HTG.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
1
Aasamizu, B., Kato, T., Sato, S., Nakamura, Y., Kaneko, T. and Tabata, S.
Structural Analysis of a Lotus japonicus Genome. IV. Sequence
Features and Mapping of seventy-three TAC clones which cover the
7.5 Mb Regions of the Genome
DNA Res. (2003) in press
2 (bases 1 to 105384)
Sato, S.
Direct Submission
Submitted (07-MAY-2003) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu,
Chiba 297-0818, Japan (E-mail: sato@kazusa.or.jp,
URL: http://www.kazusa.or.jp/, Tel: 81-438-52-3935 (ex. 2337),
Fax: 81-438-52-3934)
Location/Qualifiers
1. 105384
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"

FEATURES
source

ORIGIN
Query Match 83.2%; Score 20.8; DB 15; Length 105384;
Best Local Similarity 91.7%; Pred. No. 4.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2
AAAAAAAAAGTCCCAATTCAGATA 25
|||||
67135 AAAAAAAAAATCCCAATTCAGATA 67112

RESULT 21
AC112657 116810 bp DNA linear PRI 29-MAR-2002
LOCUS AC112657/c
DEFINITION Homo sapiens X BAC RP11-647117 (Roswell Park Cancer Institute Human
BAC library) complete sequence.
ACCESSION AC112657
VERSION AC112657.2 GI:19807700
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 116810)
Muzny, D.M., Adame, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Aye, J.R., Banks, T., Barberia, J.,
Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burke, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnheart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabies, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunatune, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, U., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loui-seged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, G., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newson, J., Newson, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenko, S.,
Ogulu, M., Okwuon, G., Oreguine, N., Ovidio, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Piums, E., Pu, L.L.,
Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibukan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoochearl, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wang, O., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wiczysk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 116810)
Morley, K.C.

```


Query Match 83.2%; Score 20.8; DB 8; Length 116810;
 Best Local Similarity 91.7%; Pred. No. 4.7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCCAATTCAGAT 25
 Db 45773 AAAAAAAAAAGTCCCTTTTCAGAT 45750

RESULT 22
 LOCUS HSRYR7C1
 DEFINITION Human DNA sequence from clone KX-PRYR7C1 on chromosome 22,
 complete sequence.
 ACCESSION AL049760 GI:5777587
 VERSION AL049760 GI:5777587
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.
 1 (bases 1 to 119118)
 Beasley, H.
 Direct Submission
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
 Clone requests: clonerequests@sanger.ac.uk
 On Aug 26, 1999 this sequence version replaced gi:5763790.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 EMBL, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 22, constructed by the Sanger Centre Chromosome 22
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr22

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one subclone; and the assembly was confirmed by restriction digest,
 except on the rare occasion of the clone being a YAC.

Location/Qualifiers
 1. 119118
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="XX-PRYR7C1"
 1
 /note="Clone left end: XX-PRYR7C1"
 join(complement(1240..2395),
 complement(283838.2:66617.105941))
 /locus_tag="RP1-127B20.3-003"
 join(complement(2191..2395),complement(1240..1305),
 complement(283838.2:101731..102081),
 complement(283838.2:99273..99413),
 complement(283838.2:81483..81534),
 complement(283838.2:78908..78984),
 complement(283838.2:76757..76811),
 complement(283838.2:75211..75233),
 complement(283838.2:73414..73572),
 complement(283838.2:71310..71389))

FEATURES
 source
 misc_feature
 gene
 mRNA

CDS
 complement(283838.2:66617..68704)
 /locus_tag="RP1-127B20.3-003"
 join(complement(2191..2244),complement(1240..1305),
 complement(283838.2:105849..105941),
 complement(283838.2:101731..102081),
 complement(283838.2:99273..99413),
 complement(283838.2:81483..81534),
 complement(283838.2:78908..78984),
 complement(283838.2:76757..76811),
 complement(283838.2:75211..75233),
 complement(283838.2:73414..73572),
 complement(283838.2:71310..71389),
 complement(283838.2:68537..68704))
 /locus_tag="RP1-127B20.3-003"
 /standard_name="OTTHUMP0000028969"
 /codon_start=1
 /protein_id="CAI23608.1"
 /db_xref="GI:56203943"
 /db_xref="GOA:Q5TFL2"
 /db_xref="InterPro:IPR001965"
 /db_xref="UniProt/TrEMBL:Q5TFL2"
 /translation="MELQSRPALAVELARHNGDILKQLHRSOPRIALSPKQALGT
 ITAVPTGPOVSLQRLAGAAVLPOVPEKTLIPSLPVGQRDPKPPKQKAT
 IVSVNPSPALPTANNVSHVPAQSPQALAPALAPLPSAGVAVALITSPENA
 AAMAPSTAVSVSDGIKNQPLISADNKVITIIPOVQCPBSTAERSPTPEBSOGAO
 ATKKKEDRPPTQENPEKIFAVNALGLVTTEHBEIQRKQKRRKSTANPAYSGILE
 TTRKLASNYLNPLPLTRANEDPCWKEITIHDECAICKRGANLQPCGTCRGATLQ
 SLIEPPLKAPKGVWCPCRCQOKLAKGKGVMTGLAVHSYSSIDRLRALRLIQG
 EQLLVTTMTTSPAPLAPMTKPSVAATHPTVQHPOGHN"
 2593..2608
 /note="Single clone region"
 2647..2669
 /note="Single clone region"
 2719..2726
 /note="Single clone region"
 50597
 /note="Clone right_end: RP4-753M9"
 73053..73133
 /note="Tandem repeat. Single clone region"
 73134
 /note="Tandem repeat. Single clone region"
 /note="Tandem repeat. Forced join. Forced join in tandem
 repeat. The repeat may not be fully represented."
 73135..73201
 /note="Tandem repeat. Single clone region"
 91746..92534
 /locus_tag="AL049760.5-005"
 /pseudo
 91746..92534
 /locus_tag="AL049760.5-005"
 /pseudo
 /codon_start=1
 119019
 /note="Clone left_end: CTA-217C2"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 8; Length 119118;
 Best Local Similarity 91.7%; Pred. No. 4.6e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
 Db 96499 AAAAAAAAAATTCCAATTCGAT 96522

RESULT 23
 LOCUS AC132412
 DEFINITION Mus musculus BAC clone RP23-355N5 from chromosome 6, complete
 sequence.
 AC132412
 AC132412
 AC132412.2 GI:48926774
 HTG.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathu; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 171747)
AUTHORS Meyer, R., Levy, A., Bielicki, L., and Haakenson, W.
TITLE The sequence of Mus musculus BAC clone RP23-355N5
JOURNAL Unpublished (2001)
AUTHORS 2 (bases 1 to 171747)
TITLE McPherson, J.D. and Waterston, R.H.
JOURNAL Direct Submission
REFERENCE 3 (bases 1 to 171747)
AUTHORS Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
TITLE Parkway, St. Louis, MO 63108, USA
JOURNAL 4 (bases 1 to 171747)
AUTHORS Wilson, R.K.
TITLE Direct Submission
REFERENCE Submitted (18-JUN-2004) Genome Sequencing Center, 4444 Forest Park
AUTHORS Parkway, St. Louis, MO 63108, USA
JOURNAL 4 (bases 1 to 171747)
AUTHORS Wilson, R.K.
TITLE Direct Submission
REFERENCE Submitted (10-JUL-2004) Genome Sequencing Center, Washington
AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
On Jun 18, 2004 this sequence version replaced gi:22657906.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BA0355N05

NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30) ; an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
and Minko Tateno in the laboratory of Pieter de Jong
(http://www.chori.org) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (http://www.resgen.com) or
Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

Location/Qualifiers
1..171747
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="6"
/map="6"
/clone="RP23-355N5"
/clone_1ib="RPCI-23"
184..343
/rpt_family="Alu"
repeat_region
384..467
/rpt_family="B4"
repeat_region

repeat_region 468..604
/rpt_family="Alu"
repeat_region 616..758
/rpt_family="Alu"
repeat_region 770..936
/rpt_family="B2"
repeat_region 954..1112
/rpt_family="B2"
repeat_region 1150..1273
/rpt_family="Alu"
repeat_region 1986..2036
/rpt_family="ID"
repeat_region 2077..2165
/rpt_family="MER2_type"
repeat_region 2141..2229
/rpt_family="MER2_type"
repeat_region 2231..2455
/rpt_family="MER2_type"
repeat_region 2478..2555
/rpt_family="ID"
repeat_region 2598..2697
/rpt_family="Alu"
repeat_region 2726..2852
/rpt_family="B4"
repeat_region 2864..2956
/rpt_family="MER2_type"
repeat_region 3071..3261
/rpt_family="B2"
repeat_region 3077..3148
/product="tRNA-Ser"
/note="Likely pseudogene (HMM Sc=38.08 / Sec struct
Sc=10.52)"
repeat_region 3511..3702
/rpt_family="B2"
repeat_region 3972..4112
/rpt_family="Alu"
repeat_region 4191..4260
/rpt_family="L2"
repeat_region 4684..4903
/rpt_family="B2"
repeat_region 4969..5112
/rpt_family="Alu"
repeat_region 5131..5308
/rpt_family="B2"
repeat_region 5325..5446
/rpt_family="Alu"
repeat_region 5345..5468
/rpt_family="MalR"
repeat_region 5484..5583
/rpt_family="B4"
repeat_region 5737..5903
/rpt_family="MalR"
repeat_region 6059..6122
/rpt_family="MalR"
repeat_region 6152..6301
/rpt_family="Alu"
repeat_region 6369..6517
/rpt_family="MalR"
repeat_region 6692..6750
/rpt_family="B4"
repeat_region 6817..6969
/rpt_family="B4"
repeat_region 6973..7022
/rpt_family="Alu"
repeat_region 7023..7159
/rpt_family="Alu"
repeat_region 7160..7220
/rpt_family="Alu"
repeat_region 7266..7324
/rpt_family="ID"
repeat_region 7701..7888
/rpt_family="B2"
repeat_region 8114..8285

```

repeat_region      /rpt_family="ERV"
                   8323..8497
                   /rpt_family="B2"
tRNA               /product="tRNA-Ser"
                   /note="Likely pseudogene (HMM Sc=29.61 / Sec struct
                   Sc=6.89)"
repeat_region      8511..8601
                   /rpt_family="B2"
repeat_region      8706..8776
                   /rpt_family="Alu"
repeat_region      9028..9099
                   /rpt_family="ID"
repeat_region      9130..9313
                   /rpt_family="B2"
tRNA               9136..9208
                   /product="tRNA-Ser"
                   /note="Likely pseudogene (HMM Sc=36.20 / Sec struct
                   Sc=12.20)"
repeat_region      9443..9614
                   /rpt_family="B2"
repeat_region      9595..9803
                   /rpt_family="B4"
repeat_region      9819..9860
                   /rpt_family="B4"
repeat_region      9861..10315
                   /rpt_family="RMR6B"
repeat_region      10316..10485
                   /rpt_family="B4"
repeat_region      10534..10675
                   /rpt_family="Alu"
repeat_region      10893..11033
                   /rpt_family="Alu"
repeat_region      12385..12516
                   /rpt_family="Alu"
repeat_region      12567..12694
                   /rpt_family="Alu"
repeat_region      12760..12868
                   /rpt_family="Alu"
repeat_region      13131..13315
                   /rpt_family="B2"
repeat_region      14191..14303
                   /rpt_family="Alu"
repeat_region      14304..14433
                   /rpt_family="L1"
repeat_region      14434..14613
                   /rpt_family="B2"

```

```

Query Match      83.2%; Score 20.8; DB 9; Length 171747;
Best Local Similarity 91.7%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 95097 AAAAAAAAAATCACAATTCAGAT 95120

```

```

RESULT 24
BX323037/c      173818 bp      DNA      linear      VRT 27-JUL-2004
LOCUS          Zebrafish DNA sequence from clone DKEYP-886 in linkage group 24,
DEFINITION     complete sequence.
ACCESSION      BX323037
VERSION        BX323037.10 GI:50724747
KEYWORDS       HTG.
SOURCE         Danio rerio (zebrafish)
ORGANISM       Danio rerio
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
               Cypriniformes; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 173818)
AUTHORS       Clark,S
TITLE         Direct Submission

```

JOURNAL
Submitted (27-JUL-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 27, 2004 this sequence version replaced gi:50057776.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Bm, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmark.shtml DKEYP-886 is from a Zebrafish BAC library VECTOR: pindigobac-5.

```

FEATURES
source          1..173818
                /organism="Danio rerio"
                /mol_type="genomic DNA"
                /db_xref="taxon:7955"
                /clone="DKEYP-886"
                /clone_1ib="Daniokeypiloc"

```

ORIGIN

```

Query Match      83.2%; Score 20.8; DB 5; Length 173818;
Best Local Similarity 91.7%; Pred. No. 4.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 76143 AAAAAAAAAAGTCAATTCAGTT 76120

```

```

RESULT 25
AC148789        174224 bp      DNA      linear      HTG 29-APR-2004
LOCUS          Otollemur garnettii clone CH256-28019, WORKING DRAFT SEQUENCE, 7
DEFINITION     Ordered pieces.
ACCESSION      AC148789
VERSION        AC148789.2 GI:46849631
KEYWORDS       HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE         Otollemur garnettii (small-eared galago)
ORGANISM       Otollemur garnettii
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Strepsirrhini;
               Galagonidae; Otollemur.
REFERENCE      1 (bases 1 to 174224)
AUTHORS       Antoneillis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,

```

Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B., Coleman,H., Daki,N., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Iodl,J.R., Jones,C., Karlins,E., Kim,H., Kwong,P., Latic,P., Larson,S., Lee-Jin,S.-Q., Legaspi,R., Maduro,P., Martic,P., Maslous,E.H., Masiello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C., Portnoy,M.E., Prasad,A., Puri,O., Reddi-Dugue,N., Schandler,K., Schuler,M.G., Shah,K., Sison,C., Stanitip,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Weheby,K.D., Young,A. and Green,E.D.
 NISC Comparative Sequencing Initiative
 2 (bases: 1 to 174224)
 Green,E.D.
 Direct Submission
 Submitted (07-APR-2004) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 3 (bases 1 to 174224)
 Green,E.D.
 Direct Submission
 Submitted (29-APR-2004) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 On Apr 29, 2004 this sequence version replaced gi:46250762.

Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_zoo@hgrl.nih.gov
 Project Information
 Center project name: gpc
 Center clone name: 028D19

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 173255 bases at least Q40
 Consensus quality: 173482 bases at least Q30
 Consensus quality: 173577 bases at least Q20
 Insert size: 175000; agarose-fp
 Insert size: 173624; sum-of-contigs
 Quality coverage: 9.95x in Q20 bases; agarose-fp
 Quality coverage: 10.02x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
 1 20218: contig of 20218 bp in length
 20219: gap of 24198 bp in length
 20319: gap of unknown length
 44516: gap of unknown length
 44517: gap of unknown length
 74217: contig of 29601 bp in length
 74218: gap of unknown length
 74318: contig of 6141 bp in length
 80458: gap of unknown length
 80459

FEATURES
 source
 * 80559 88789: contig of 8231 bp in length
 * 88790 88889: gap of unknown length
 * 88890 93362: contig of 4473 bp in length
 * 93363 93462: gap of unknown length
 * 93463 174224: contig of 80762 bp in length.
 Location/Qualifiers
 1..174224
 /organism="Oreolemur garnettii"
 /mol_type="genomic DNA"
 /db_xref="taxon:30611"
 /clone="CH256-28D19"
 /clone_1ib="CH256"
 /note="BAC resource: <http://bacpac.chori.org/>"

misc_feature
 1..20218
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 20219..20318
 /estimated_length=unknown
 20319..44516
 /note="assembly_fragment"
 44517..44616
 /estimated_length=unknown
 44617..74217
 /note="assembly_fragment"
 74218..74317
 /estimated_length=unknown
 74318..80458
 /note="assembly_fragment"
 80459..80558
 /estimated_length=unknown
 80559..88789
 /note="assembly_fragment"
 88790..88889
 /estimated_length=unknown
 88890..93362
 /note="assembly_fragment"
 93363..93462
 /estimated_length=unknown
 93463..174224
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 174224;
 Best Local Similarity 91.7%; Pred. No. 4.1e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
 Db 62316 AAAAAAAAAAGTCCCAATTCAGAT 62339

RESULT 26

AC118957 175991 bp DNA linear HTG 20-NOV-2002
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-436C18, WORKING DRAFT SEQUENCE, 3
 unnumbered pieces.
 AC118957
 AC118957.12 GI:25137755
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
 1 (bases 1 to 175991)
 Muzny,D.,Maric,M.,Metzger,M.,Lee,A.,Abramson,S.,Adams,C.,Alder,J.,Allen,C.,Allen,H.,Alsbrooks,S.,Amin,A.,Anguiano,D.,Anyalebech,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,Baldwin,D.,Bandaranaike,D.,Barber,M.,Barnstead,M.,Benahmed,F.,Biswal,K.,Blair,J.,Blankenburg,K.,Blythe,P.,Brown,M.,

Bryan, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davis, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Huik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louieged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minda, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelimeh, O., Okwuonu, G., Olarnunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, D., Shervatsbayan, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K., Valae, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, X., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Unpublished
2 (bases 1 to 175991)
Direct Submission
Submitted (22-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 175991)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23196037.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center
Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVDX
Center clone name: CH230-436C18
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 166257 bases at least Q40
Consensus quality: 168339 bases at least Q30
Consensus quality: 169676 bases at least Q20
Estimated insert size: 171657; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3282: contig of 3282 bp in length
* 3283 3282: gap of unknown length
* 3383 174786: contig of 174786 bp in length
* 174787 174886: gap of unknown length
* 174887 175991: contig of 1105 bp in length.
* Location/Qualifiers
1..175991
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-436C18"
1..1276
/note="wgs_end-extension
clone_end:T7"
1829..2693
/note="clone_boundary
clone_end:T7
site:
end_sequence: B2203345"
2166..3282
/note="wgs_contig"
3283..3382
/estimated_length=unknown
3383..5160
/note="wgs_contig"
169986..170901
/note="clone_boundary
clone_end:Sp6
site:
end_sequence: B2203346"
174787..174886
/estimated_length=unknown
gap
Query Match 83.2%; Score 20.8; DB 14; Length 175991;
Best local Similarity 91.7%; Pred. No. 4,1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCCAATTCAGT 24
Db 130488 AAAAAAAAAAGTCCCAATTCAGT 130511
RESULT 27
AC022034 195768 bp DNA linear PRI 01-JAN-2002
LOCUS Homo sapiens chromosome 8, clone RP11-400K9, complete sequence.
DEFINITION AC022034
ACCESSION AC022034.7 GI:17426331
VERSION


```

KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 195768)
Homo sapiens chromosome 8, clone RP11-400K9
Unpublished
2 (bases 1 to 195768)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fennestor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Mcdonald, P., Margulis, N., Mckwan, P., McGurk, A., McKernan, K.,
Mcneuers, R., Meldrum, J., Menes, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tjirell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 195768)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Gand, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
McClean, C., Mcdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., Mckwan, P., McKernan, K., Mcneuers, R., Meldrum, J.,
Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Punthang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All Dec 8, 2001 this sequence version replaced g1:1786392.
All repeats were identified using RepeatMasker:
http://fdd.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2101
Center clone name: 400_K_9
----- Location/Qualifiers
1. 195768
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-400K9"
/clone_1lb="RP11 Human Male BAC"
914. 1047
/rpt_family="MIR3"
1634. 1928
/rpt_family="ALUSg"
complement(5008. 5192)
/rpt_family="MERSA"
complement(5686. 5778)
/rpt_family="MIR"
complement(5817. 5972)
/rpt_family="MIR3"
6807. 6894
/rpt_family=" (TG)n"
7044. 7299
/rpt_family="MIR"
complement(8103. 8247)
/rpt_family="FLAM C"
complement(9261. 9553)
/rpt_family="ALUSg"
10325. 10358
/rpt_family="AT_rich"
10794. 10901
/rpt_family="MER96B"
11864. 11884
/rpt_family="AT_rich"
12585. 12658
/rpt_family="MER113"
complement(12696. 12900)
/rpt_family="L1MC5"
12901. 13207
/rpt_family="ALUSx"
complement(13208. 13252)
/rpt_family="L1MC5"
13321. 13718
/rpt_family="MER6A"
/rpt_family="L1MD1"
complement(13719. 14021)
/rpt_family="ALUSx"
14022. 14091
/rpt_family="MER6A"
complement(14192. 14356)
/rpt_family="L1MC5"
complement(14418. 15549)
/rpt_family="L1MC5"
complement(15545. 15682)
/rpt_family="L1MD1"
complement(15685. 17136)
/rpt_family="L1MD2"
17137. 17440
/rpt_family="ALUSb"
complement(17441. 18005)
/rpt_family="L1MD2"
complement(18002. 19359)
/rpt_family="L1MD2"
19384. 19509
/rpt_family="MIR"
complement(19606. 19983)
/rpt_family="MER113"
19990. 20204
/rpt_family="L1MC4a"
20205. 20238
/rpt_family=" (CA)n"
20396. 20601
/rpt_family="L1MDa"
complement(20602. 20753)
/rpt_family="TRH1C"
complement(20756. 21978)
/rpt_family="MSRA-inc"
21979. 22016

```

FEATURES

```

source
Location/Qualifiers
1. 195768
/organism="Homo sapiens"

```



```

gap      41590..41639
         /estimated_length=50
gap      48630..48917
         /estimated_length=288
gap      53842..53891
         /estimated_length=50
gap      122065..122114
         /estimated_length=50
gap      131574..131623
         /estimated_length=50
gap      161012..161061
         /estimated_length=50
gap      168784..168833
         /estimated_length=50
gap      182847..182896
         /estimated_length=50
gap      189168..189260
         /estimated_length=93
gap      201740..201789
         /estimated_length=50
gap      217685..217734
         /estimated_length=50
         /estimated_length=50

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 225678;
Best Local Similarity 91.7%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db      148822 AAAAAAAAAAGTCCCAATTCAGAT 148845

RESULT 30
LOCUS   AC166947/c
DEFINITION Bos taurus clone CH240-189F6, *** SEQUENCING IN PROGRESS ***, 24
ACCESSION AC166947
VERSION   AC166947.2 GI:72535224
KEYWORDS  HTG; HTGS PHASE1.
SOURCE    Bos taurus (cow)
ORGANISM  Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
          Pecora; Bovidae; Bovinae; Bos.
          1 (bases 1 to 86584)
          Muzny,D., Adams,C., Agbat II,O., Allen,C., Alsbrooke,S., Archer,P.,
          Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
          Beraducci,A., Biswal,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
          Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
          Chacoe,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
          Cheng,M.-T., Chu,D., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
          Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
          Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
          Donlin,I., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
          Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
          Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
          Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
          Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
          Gonzalez-Garay,M., Guevara,W., Holder,M., Haealand,W., Haeblerlen,K.,
          Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
          Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
          Hines,S., Hitchens,M., Hodgson,A., Hogue,M., Hollins,B.,
          Howell,L.T., Hulik,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
          Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,
          Keys,T., Khan,Z., King,L., Koyar,C., Kowis,A., Kowis,C., Lara,F.,
          Leal,S., Lee,K., Lee,S., Legall,F.I., Lemon,S., Lewis,Li,B.,
          Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Llyanage,D.,
          London,P., Lopez,J., Lorensuhewa,L., Lozano,R., Luk,T., Madu,R.,
          Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
          McClelland,H., McPherson,J., Mercadado,C., Metzker,M.,
          Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Munitase,M.,

```

```

Murray,D., Nazareth,L., Ngo,D., Nguyen,N., Norvig-Eaerugh,E.,
Nott,A., Nwokelelehen,O., Obregon,M., Och-Ofori,C., Oden,E.,
Okunolu,G., Okunolu,K., Parker,D., Pasternak,S., Patel,B.,
Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
Primus,E., Pu,L.-L., Puazo,M., Qian,X., Quinn,A., Quiroz,J.,
Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sama,M., Sanders,W.,
Sanibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,
Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorm,R., Thornton,R.,
Trejos,Z., Uমান,I., Vargo,C., Verdusco,D., Villaneda,D., Vitek,D.,
Volkov,A., Waldron,L., Walker,B., Wang,O., Wang,S., Warren,J.,
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
Wu,D., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,
Zhang,Z., Zhou,J., Weinstein,G., and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

```

```

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help.tmc.edu

----- Project Information
Center project name: FLBY
Center clone name: CH240-189F6

----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 9386 bases at least Q40
Consensus quality: 99091 bases at least Q30
Consensus quality: 103084 bases at least Q20
Estimated insert size: 111635; sum-of-contigs estimation
Estimated insert size: 79065; agarose-fp estimation
Quality coverage: 2x in Q20 bases; agarose-fp estimation
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      4923: contig of 4923 bp in length
*
4924      5023: gap of unknown length
*
5024      8840: contig of 3817 bp in length
*
8841      8940: gap of unknown length
*
8941      11341: contig of 2401 bp in length
*
11342      11441: gap of unknown length
*
11442      13581: contig of 2140 bp in length
*
13582      13681: gap of unknown length
*
13682      16394: contig of 2713 bp in length
*
16395      16494: gap of unknown length
*
16495      18726: contig of 2132 bp in length
*
18727      21964: gap of unknown length
*
21964: contig of 3238 bp in length

```

```

* 21965 22064: gap of unknown length
* 22065 27338: contig of 5274 bp in length
* 27339 27438: gap of unknown length
* 30344 30344: contig of 2906 bp in length
* 30345 30444: gap of unknown length
* 34785 34785: contig of 4341 bp in length
* 34885 34885: gap of unknown length
* 37443 37443: contig of 2558 bp in length
* 37444 37543: gap of unknown length
* 40139 40139: contig of 2596 bp in length
* 40240 40239: gap of unknown length
* 43014 43014: contig of 2775 bp in length
* 43115 43115: gap of unknown length
* 45357 45356: contig of 2242 bp in length
* 45457 45456: gap of unknown length
* 51222 51221: contig of 5765 bp in length
* 51322 53649: gap of unknown length
* 53650 53749: contig of 2328 bp in length
* 53750 56745: gap of unknown length
* 56746 56845: contig of 2996 bp in length
* 56846 61746: gap of unknown length
* 61747 61846: contig of 4901 bp in length
* 61847 64188: gap of unknown length
* 64189 64288: contig of 2342 bp in length
* 64289 70113: gap of unknown length
* 70114 70213: contig of 5825 bp in length
* 70214 72912: gap of unknown length
* 72913 73012: contig of 2699 bp in length
* 73013 76098: gap of unknown length
* 76099 76198: contig of 3086 bp in length
* 76199 84389: gap of unknown length
* 84390 84489: contig of 8191 bp in length
* 84490 86584: gap of unknown length
* 86584 86584: contig of 2095 bp in length.
* 86584 86584: Location/Qualifiers
* 1. 86584
* /organism="Bos taurus"
* /mol_type="genomic DNA"
* /db_xref="taxon:9913"
* /clone="CH240-189F6"
* 4924. .5023
* /estimated_length=unknown
* 8841. .8940
* /estimated_length=unknown
* 11342. .1141
* /estimated_length=unknown
* 13582. .13681
* /estimated_length=unknown
* 16395. .16494
* /estimated_length=unknown
* 18627. .18726
* /estimated_length=unknown
* 21965. .22064
* /estimated_length=unknown
* 27339. .27438
* /estimated_length=unknown
* 30345. .30444
* /estimated_length=unknown
* 34786. .34885
* /estimated_length=unknown
* 37444. .37543
* /estimated_length=unknown
* 40140. .40239
* /estimated_length=unknown
* 43015. .43114
* /estimated_length=unknown
* 45357. .45456
* /estimated_length=unknown
* 51222. .51321
* /estimated_length=unknown
* 53650. .53749
* /estimated_length=unknown
* 56746. .56845

```

```

ORIGIN
Query Match      81.6%; Score 20.4; DB 14; Length 86584;
Best Local Similarity 95.5%; Pred. No. 7.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAG 22
Db 62236 AAAAAAAAAAGTCCCAATTCAG 62215

RESULT 31
HS298E2/C
LOCUS      140741 bp      DNA      linear      PRI 29-MAY-2002
DEFINITION Homo sapiens chromosome 9 BAC Rpl1-298E2, complete sequence.
ACCESSION  AL163192
VERSION     AL163192.6  GI:19572640
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.
            1 (bases 1 to 140741)
            Scharfe,M., Conrad,A., Hornischer,K., Loehnert,T.H., Thies,S. and
            Bloecker,H.
            Direct Submission
            Submitted (05-APR-2000) GBF, Dept. of Genome Analysis, Mascheroder
            Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@bf.de
            On Mar 21, 2002 this sequence version replaced gi:18873645.
            PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
            + Analysis and annotation were performed with the automatic
            + 'first-pass' annotation and submission tool
            + 'AnnotMitter' (Hornischer & Bloecker).
            + Programs used by 'AnnotMitter':
            + ++++++
            + > GeneFinder (Green), Vers. 084
            + > Organism: human
            + > Genscan (Burge & Karlin), Vers. 1.0
            + > Used matrix: vertebrate; Minimum score: 0
            + > Grail (Xu et al.), Vers. 1.3
            + > Organism: human
            + > Mzef (Zhang)
            + > Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
            + & Skolnick)
            + > Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
            + BLASTN 2.0.14 (Altschul et al.)
            + > Database(s): * RepBase: ALU (human), released 22-DEC-1995 .
            + * RepBase: L1 (primate), released 22-DEC-1995 .
            + * RepBase: MIR (primate), released 22-DEC-1995 .
            + * RepBase: MIR2 (primate), released 22-DEC-1995 .
            + * RepBase: THE (primate), released 22-DEC-1995 .
            + Minimum identity: 70 %;
            + > 'ESTs': BLASTN 2.0.14 (Altschul et al.)
            + > Database(s): * emb1 (EST, human), released -DEC- .
            + * emb1 (EST, other), released -DEC- .
            + (EST), Vers. 67+ (01-JAN-1970) . Using sequence with masked

```

```

repeats
. Minimum score: 60; Minimum identity: 70 %;
> 'Tandem Repeats': GDE 2.2 option 'tandem'
. Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
> 'Treat N's as mismatches? YES; Allow uniform consensus? NO >
> 'Inverted Repeats': GDE 2.2 option 'inverted'
> 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'Cpg
Islands': GDE 2.2 option 'cpg'
. Cpg Island region size 100 bp;
. Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS scan':
e-PCR (Schuler)
. Maxin: 50; Number of mismatches allowed: 0; Word size: 7
STS database: 'dbsts markers'
> 'cRNA Scan': cRNAcan-SS (Lowe & Eddy), Vers. 1.11 All
annotations in this database entry are developed by computational
tools. It is therefore not explicitly noted in the feature lines
that evidence is not experimental.
Mapping was performed at The Sanger Centre
(cf. http://www.sanger.ac.uk/HGP/Chr9)
Mapping information is available via
http://webace.sanger.ac.uk/cgi-bin/display?db=acedb9&grep=298E2
----- Genome Center
Center: GBF, Braunschweig
Center code: GBF
Web site: http://genome.gbf.de/
Contact: info.genome@gbf.de
----- Project Information
Center project name:
Center clone name: h51
----- Summary Statistics
Sequencing vector: ##;
Chemistry: Dye-terminator-amerham: 71% of reads
Chemistry: Dye-primer-amerham: 29% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 0 bases at least Q40
Consensus quality: 0 bases at least Q30
Consensus quality: 0 bases at least Q20
Estimated insert size: ##; agarose-fp estimation
Estimated insert size: 140741; sum-of-ctrls estimation
-----
Location/Qualifiers
source
1..140741
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-298B2"
/clone_1lb="RPC1-11.2"
1..140741
/feature="assembly_fragment-clone end:T7-vector_side:left
assembly_fragment-clone_end:SP6-vector_side:right"
complement(3..1429)
/note="88% identity: matches 2180..3616 of consensus"
/rpt_family="L1"
complement(1459..1733)
/note="85% identity: matches 427..698 of consensus"
/rpt_family="ALU"
complement(1817..2863)
/note="88% identity: matches 617..1663 of consensus"
/rpt_family="L1"
2484..2572
/note="MZEF prediction, score = 0.654"
complement(2848..3135)
/note="GRAIL, score = 50%, comment = good shadow"
complement(2914..2984)
/note="85% identity: matches 951..1021 of consensus"
/rpt_family="L1"
complement(2954..3643)

```

```

/note="89% identity: matches 717..1404 of consensus"
/rpt_family="L1"
3619..3659
/note="XPOUND prediction, score = 0.288"
complement(3696..3924)
/note="87% identity: matches 445..664 of consensus"
/rpt_family="L1"
complement(3976..4364)
/note="84% identity: matches 1..393 of consensus"
/rpt_family="L1"
4029..4189
/note="GRAIL, score = 50%, comment = good"
4946..5047
/note="GRAIL, score = 98%, comment = excellent
MZEF prediction, score = 0.707"
5115..5208
/note="GRAIL, score = 64%, comment = good"
5293..5449
/standard_name="TIGR-A004T44 (D12S1978), Map: 74.9, Homo
sapiens"
/note="GenBank Accession Number: G26435"
5335..5439
/standard_name="SHGC-130022 (D11S4975), Map: 9, Homo
sapiens"
/note="GenBank Accession Number: G59473"
6650..6683
/note="XPOUND prediction, score = 0.552"
6850..6915
/note="homology = 68.2%, counts = 33"
/rpt_family="CA repeat"
/rpt_type="TANDEM
6850..6864
/note="CA repeat"
8245..8260
/note="XPOUND prediction, score = 0.250"
complement(9803..9915)
/note="GENSCAN prediction, score = 6.82
GRAIL, score = 81%, comment = excellent"
complement(9887..9915)
/note="XPOUND prediction, score = 0.219"
10147..10160
/note="TTC repeat"
complement(10560..10590)
/note="XPOUND prediction, score = 0.298"
complement(10626..10694)
/note="XPOUND prediction, score = 0.236"
10768..10784
/note="TG repeat"
complement(11362..11527)
/note="80% identity: matches 67..232 of consensus"
/rpt_family="ALU"
complement(11491..11521)
/note="XPOUND prediction, score = 0.539"
complement(11534..11588)
/note="94% identity: matches 2..56 of consensus"
/rpt_family="ALU"
complement(11602..11606)
/note="XPOUND prediction, score = 0.209"
11649..11661
/note="TG repeat"
complement(13184..13342)
/note="GRAIL, score = 90%, comment = excellent"
14140..14308
/note="87% identity: matches 245..413 of consensus"
/rpt_family="L1"
complement(14140..14308)
/note="92% identity: matches 86..253 of consensus"
/rpt_family="ALU"
14308..14403
/note="91% identity: matches 431..526 of consensus"
/rpt_family="L1"
complement(14308..14403)
/note="91% identity: matches 292..387 of consensus"

```

```

exon          /rpt_family="ALU"
              14581..14541
              /note="GRATL, score = 57%, comment = good shadow"
repeat_region 15850..15924
              /note="IR1, 85% complementary to IR1' (15959..16033)"
              /rpt_type=INVERTED
repeat_region 15959..16033

Query Match      81.6%; Score 20.4; DB 8; Length 140741;
Best Local Similarity 95.5%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 AAAAAAAAAAGTCCCAATTGAG 22
        |||
Db      92636 AAAAAAAAAAGTCCCAATTGAG 92615

RESULT 32
AC140436
LOCUS
DEFINITION Mus musculus BAC clone RP23-8502 from chromosome 9, complete
sequence.
ACCESSION AC140436
VERSION AC140436.3 GI:32964985
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muroidea; Muridae; Murinae; Mus.
          1 (bases 1 to 142560)
          Bauer, H.
REFERENCE
AUTHORS The sequence of Mus musculus BAC clone RP23-8502
TITLE Unpublished (2001)
JOURNAL 2 (bases 1 to 142560)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
AUTHORS 3 (bases 1 to 142560)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS 4 (bases 1 to 142560)
TITLE Submitted (05-JUN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS 5 (bases 1 to 142560)
TITLE Submitted (05-JUN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS 6 (bases 1 to 142560)
TITLE Submitted (25-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL On Jul 18, 2003 this sequence version replaced gi:31416111.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_BA0085002

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RP23-8502 BAC library has been constructed by Kazutoyo Osewaga and Minako Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC124352.

FEATURES

source

```

Location/Qualifiers
1..142560
/organism="Mus musculus"
/mol_type="genomic DNA"
/DB_xref="taxon:10090"
/chromosome="9"
/map="9"
/clone="RP23-8502"
/clone_1b="RP23-8502"
1..1257
/rpt_family="L1"
repeat_region 1526..1683
/rpt_family="B2"
repeat_region 2257..2379
/rpt_family="ALU"
repeat_region 2764..2892
/rpt_family="B4"
repeat_region 3912..4316
/rpt_family="L1"
repeat_region 4439..4511
/rpt_family="TD"
repeat_region 5965..6130
/rpt_family="B4"
repeat_region 6496..6690
/rpt_family="B2"
repeat_region 7031..7351
/rpt_family="MALR"
repeat_region 7417..7470
/rpt_family="L2"
repeat_region 7933..8055
/rpt_family="ALU"
repeat_region 8221..8491
/rpt_family="MALR"
repeat_region 9284..9655
/rpt_family="MALR"
repeat_region 16058..16463
/rpt_family="L1"
repeat_region 16782..17157
/rpt_family="MALR"
repeat_region 17162..18200
/rpt_family="L1"
repeat_region 18445..18608
/rpt_family="L1"
repeat_region 18639..18914
/rpt_family="L1"
repeat_region 18917..19828
/rpt_family="L1"
repeat_region 20381..20683

```

```

/rpt_family="RMER1B"
repeat_region 20884..21091
/rpt_family="L1"
repeat_region 21087..22368
/rpt_family="L1"
repeat_region 23790..24106
/rpt_family="L1"
repeat_region 24093..24697
/rpt_family="L1"
repeat_region 24727..24993
/rpt_family="L1"
repeat_region 25440..26188
/rpt_family="L1"
repeat_region 27643..27818
/rpt_family="B2"
repeat_region 27838..27969
/rpt_family="L1"
repeat_region 28447..28562
/rpt_family="Alu"
repeat_region 29536..29690
/rpt_family="B4"
repeat_region 29782..29852
/rpt_family="MIR"
repeat_region 31383..31509
/rpt_family="MER1_type"
repeat_region 32207..32270
/rpt_family="MIR"
repeat_region 33288..33498
/rpt_family="MIR"
repeat_region 34268..34756
/rpt_family="L1"
repeat_region 35616..36144
/rpt_family="L1"
repeat_region 36169..36434
/rpt_family="MER99"
repeat_region 36463..36665
/rpt_family="B2"
repeat_region 36734..36820
/rpt_family="MER99"
repeat_region 36938..37527
/rpt_family="L1"
repeat_region 39286..39468
/rpt_family="MAlR"
repeat_region 40064..40171
/rpt_family="ACHobo"
repeat_region 41095..41430
/rpt_family="MAlR"
repeat_region 41467..41621
/rpt_family="Alu"
repeat_region 41990..42262
/rpt_family="L1"
repeat_region 42317..42396
/rpt_family="B2"
repeat_region 42443..42801
/rpt_family="RMER1S"
repeat_region 43035..43129
/rpt_family="Alu"
repeat_region 43189..43467
/rpt_family="B4"
repeat_region 43546..43606
/rpt_family="ERY1"
repeat_region 44341..44731
/rpt_family="MAlR"
repeat_region 45293..45460
/rpt_family="MIR"
repeat_region 45545..45769
/rpt_family="MAlR"
repeat_region 46017..47178
/rpt_family="L1"
repeat_region 47262..48569
/rpt_family="L1"
repeat_region 49541..49751
/rpt_family="B2"

```

```

Query Match      81.6%; Score 20.4; DB 9; Length 142560;
Best Local Similarity 95.5%; Pred. No. 6,1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTCCATTCAG 22
Db      51193 AAAAAAAAAAGTCCATTCAG 51214

RESULT 33
AC158834/c
LOCUS 8834
DEFINITION
ACCESSION
AC158834.1 GI:61740706
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Muzny,D., Adams,C., Agbai II,O., Allen,C., Albrooks,S., Archer,P.,
Arrondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
Berducci,A., Biewalo,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
Chacko,J., Chahrour,M., Chavez,D., Chen,R., Chen,G., Chen,R.,
Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyte,M., Cree,A.,
Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
Dziunda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
Gonzalez-Garay,M., Guvera,W., Holder,M., Haland,M., Haeblerlein,K.,
Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
Hines,S., Hitchens,M., Hodgson,A., Hogue,M., Hollins,B.,
Howell,L.T., Huliy,S., Hume,J., Iino,K., Jackson,A., Jackson,L.,
Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,
Keys,T., Khan,Z., King,L., Kovar,C., Kowals,A., Kowals,C., Lara,F.,
Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,
London,P., Lopez,J., Lorensuhera,L., Lozano,R., Luk,T., Madu,R.,
Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
McClelland,H., McPherson,J., Mercedao,C., Metzker,M.,
Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Muntada,M.,
Murray,D., Nazareth,L., Ngo,D., Nguyen,N., Norwig-Bastush,E.,
Nott,A., Nwokedeme,O., Obregon,M., Och-Ofori,C., Odeh,E.,
Okwumodu,G., Okwumodu,K., Parker,D., Pasternak,S., Patel,B.,
Patel,V., Paul,H., Perez,A., Perez,L., Petrovino,J., Pham,T.,
Primus,E., Pu,L.-L., Puzro,M., Qin,X., Quinn,A., Quiroz,J.,
Rabata,D., Rachlin,E., Reigis,R., Ren,Y., Reuter,M., Richards,S.,
Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,
Sanctibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,
Stavrek,A., Teylor,E., Taylor,C., Thomas,N., Thoren,R.,
Trejos,Z., Uemami,K., Vargo,C., Vezduco,D., Villaseca,D., Vink,D.,
Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
Wu,J., Yakub,S., Yan,K., Yan,Y., Yu,F., Zhang,J., Zhang,L.,
Zhang,Z., Zhou,J., Weinstein,G. and Gibbs,R.

Direct Submission
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 156860)
Worley,K.C.
Direct Submission
Submitted (24-MAR-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```


JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished (2001)
2. (bases 1 to 169938)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (28-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3. (bases 1 to 169938)
Wilson, R.K.
Direct Submission
Submitted (07-MAR-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4. (bases 1 to 169938)
Wilson, R.K.
Direct Submission
Submitted (15-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Mar '7, 2004 this sequence version replaced gi:29336270.

Center: Washington University Genome Sequencing Center
 Center code: WUSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 Summary Statistics
 Center project name: M_BB0531P16

NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPC1-2 BAC library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC124352.

FEATURES

Sources

Location/Qualifiers
1. .169938

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="9"
/map="9"
/clone="RP24-531P16"
/clone_1fb="RPC1-24"
2702.._2792
/note="Sequence derived from one plasmid subclone."
misc_feature
22325..23317
/note="CpG island (tGC=58.5, o/e=0.84, #CpGs=70)"
26694..30023
/note="CpG island (tGC=64.6, o/e=0.68, #CpGs=97)"
misc_feature

```

ORIGIN

Query Match	81.6%	Score 20.4	DB 9	Length 169938
Best Local Similarity	95.5%	Pred. No. 5.8e+02		
Matches 21; Conservative	0	Mismatches 1	Indels 0	Gaps 0

```

QY      1  AAAAAAAAAAGTCCCAATTCAg  22
          |||||
Db      128645  AAAAAAAAAAGTCCCAATTCAG  128666

```

RESULT	35
AC124352	
LOCUS	177886 bp DNA linear ROD 13-NOV-2003
DEFINITION	Mus musculus BAC clone RP24-47/C23 from chromosome 9, complete
ACCESSION	AC124352
VERSION	AC124352.
KEYWORDS	AC124352.2 GI:23592197
SOURCE	HMG.
ORGANISM	Mus musculus (house mouse)
	Mus musculus

REFERENCE 1 (bases 1 to 177886)
AUTHORS Berghoff, A., Meyer, R. and Mangiapanello, L.
TITLE The sequence of Mus musculus BAC clone RP24-477C22
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 177886)
AUTHORS Wilson, R.

JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 177886)
AUTHORS	McPherson, J.D. and Waterston, R.H

JOURNAL
Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Submitted 04-JUN-2002

AUTHORS
McPherson, J. D. and Waterston, R. H.

JOURNAL Submitted (08-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 177886)

REFERENCE	5 (bases 1 to 177886)
AUTHORS	Wilson, R.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Oct 8, 2002 this sequence version replaced gi:21426472.

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.wustl.edu
 ----- Summary Statistics -----
 Center project name: M_BB0477CC23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see: <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pletter de Jong and

coworkers (http://www.chori.org) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers

FEATURES

```
source
1..177886
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="9"
/map="9"
/clone="RP24-47C23"
/clone_lib="R2CT-24"
848..1534
/rpt_family="L1"
1552..2157
/rpt_family="L1"
4211..4277
/rpt_family="MIR"
4617..4902
/rpt_family="MALR"
10525..10989
/rpt_family="L1"
14733..14919
/rpt_family="L1"
14960..15017
/rpt_family="ERV1"
15441..15637
/rpt_family="L1"
15639..15719
/rpt_family="Alu"
15826..16060
/rpt_family="MALR"
16062..16441
/rpt_family="ERVK"
16442..17174
/rpt_family="ERVK"
20540..20919
/rpt_family="ERVK"
20926..20963
/rpt_family="MALR"
20964..21370
/rpt_family="MALR"
21371..21444
/rpt_family="MALR"
21836..22517
/rpt_family="MALR"
24364..24575
/rpt_family="RMR6A"
24769..24912
/rpt_family="B2"
25954..26062
/rpt_family="B4"
26058..26314
/rpt_family="MALR"
26315..27273
/rpt_family="MALR"
27592..28535
/rpt_family="MALR"
29078..29425
/rpt_family="L1"
29930..30083
/rpt_family="B4"
30305..30389
/rpt_family="ERV1"
37135..37292
/rpt_family="B2"
37866..37988
/rpt_family="Alu"
38373..38501
/rpt_family="B4"
```

```
repeat_region 39521..39925
/rpt_family="L1"
repeat_region 40048..40120
/rpt_family="ID"
repeat_region 41574..41739
/rpt_family="B4"
repeat_region 42105..42299
/rpt_family="B2"
repeat_region 42640..42960
/rpt_family="MALR"
repeat_region 43026..43079
/rpt_family="L2"
repeat_region 43542..43664
/rpt_family="Alu"
repeat_region 43830..44100
/rpt_family="MALR"
repeat_region 44893..45264
/rpt_family="MALR"
repeat_region 51667..52072
/rpt_family="L1"
repeat_region 52391..52766
/rpt_family="MALR"
repeat_region 52771..53809
/rpt_family="L1"
repeat_region 54054..54217
/rpt_family="L1"
repeat_region 54248..54523
/rpt_family="L1"
repeat_region 54526..55437
/rpt_family="L1"
repeat_region 55890..56292
/rpt_family="RMR1B"
repeat_region 56293..56700
/rpt_family="L1"
repeat_region 56696..57977
/rpt_family="L1"
repeat_region 59399..59715
/rpt_family="L1"
repeat_region 59702..60306
/rpt_family="L1"
repeat_region 60336..60602
/rpt_family="L1"
repeat_region 61049..61797
/rpt_family="L1"
repeat_region 63252..63427
/rpt_family="B2"
repeat_region 63447..63578
/rpt_family="L1"
repeat_region 64056..64177
/rpt_family="Alu"
repeat_region 65145..65299
/rpt_family="B4"
repeat_region 65391..65461
/rpt_family="MIR"
repeat_region 66992..67118
/rpt_family="MIR1_type"
repeat_region 67816..67879
/rpt_family="MIR"
repeat_region 68897..69107
/rpt_family="MIR"
repeat_region 69677..70365
/rpt_family="L1"
```

Query Match 81.6%; Score 20.4; DB 9; Length 177886;
Best Local Similarity 95.5%; Pred. No. 5.7e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 1 AAAAAAAAAAGTCCCAATTCAG 22
DB 86802 AAAAAAAAAAGTCCCAATTCAG 86823
```

RESULT 36

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AC157436	1	HTG, PHASE2, HTGS DRAFT.	Oryctolagus cuniculus (rabbit)	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha; Leporidae; Oryctolagus.
AC157436.2	GI:63094054			
AC157436	1	(bases 1 to 180048)		
AC157436	2	(bases 1 to 180048)		
AC157436	3	(bases 1 to 180048)		
AC157436	4	(bases 1 to 180048)		
AC157436	5	(bases 1 to 180048)		
AC157436	6	(bases 1 to 180048)		
AC157436	7	(bases 1 to 180048)		
AC157436	8	(bases 1 to 180048)		
AC157436	9	(bases 1 to 180048)		
AC157436	10	(bases 1 to 180048)		
AC157436	11	(bases 1 to 180048)		
AC157436	12	(bases 1 to 180048)		
AC157436	13	(bases 1 to 180048)		
AC157436	14	(bases 1 to 180048)		
AC157436	15	(bases 1 to 180048)		
AC157436	16	(bases 1 to 180048)		
AC157436	17	(bases 1 to 180048)		
AC157436	18	(bases 1 to 180048)		
AC157436	19	(bases 1 to 180048)		
AC157436	20	(bases 1 to 180048)		
AC157436	21	(bases 1 to 180048)		
AC157436	22	(bases 1 to 180048)		
AC157436	23	(bases 1 to 180048)		
AC157436	24	(bases 1 to 180048)		
AC157436	25	(bases 1 to 180048)		
AC157436	26	(bases 1 to 180048)		
AC157436	27	(bases 1 to 180048)		
AC157436	28	(bases 1 to 180048)		
AC157436	29	(bases 1 to 180048)		
AC157436	30	(bases 1 to 180048)		
AC157436	31	(bases 1 to 180048)		
AC157436	32	(bases 1 to 180048)		
AC157436	33	(bases 1 to 180048)		
AC157436	34	(bases 1 to 180048)		
AC157436	35	(bases 1 to 180048)		
AC157436	36	(bases 1 to 180048)		
AC157436	37	(bases 1 to 180048)		
AC157436	38	(bases 1 to 180048)		
AC157436	39	(bases 1 to 180048)		
AC157436	40	(bases 1 to 180048)		
AC157436	41	(bases 1 to 180048)		
AC157436	42	(bases 1 to 180048)		
AC157436	43	(bases 1 to 180048)		
AC157436	44	(bases 1 to 180048)		
AC157436	45	(bases 1 to 180048)		
AC157436	46	(bases 1 to 180048)		
AC157436	47	(bases 1 to 180048)		
AC157436	48	(bases 1 to 180048)		
AC157436	49	(bases 1 to 180048)		
AC157436	50	(bases 1 to 180048)		
AC157436	51	(bases 1 to 180048)		
AC157436	52	(bases 1 to 180048)		
AC157436	53	(bases 1 to 180048)		
AC157436	54	(bases 1 to 180048)		
AC157436	55	(bases 1 to 180048)		
AC157436	56	(bases 1 to 180048)		
AC157436	57	(bases 1 to 180048)		
AC157436	58	(bases 1 to 180048)		
AC157436	59	(bases 1 to 180048)		
AC157436	60	(bases 1 to 180048)		
AC157436	61	(bases 1 to 180048)		
AC157436	62	(bases 1 to 180048)		
AC157436	63	(bases 1 to 180048)		
AC157436	64	(bases 1 to 180048)		
AC157436	65	(bases 1 to 180048)		
AC157436	66	(bases 1 to 180048)		
AC157436	67	(bases 1 to 180048)		
AC157436	68	(bases 1 to 180048)		
AC157436	69	(bases 1 to 180048)		
AC157436	70	(bases 1 to 180048)		
AC157436	71			

FEATURES	SOURCE
* are represented as runs of N. The order of the pieces	
* is believed to be correct as given, however the sizes	
* of the gaps between them are based on estimates that have	
* provided by the submitter.	
* This sequence will be replaced	
* by the finished sequence as soon as it is available and	
* the accession number will be preserved.	
* 1 20929: contig of 20929 bp in length	
* 21029: gap of unknown length	
* 21030 30083: contig of 9054 bp in length	
* 30084 30183: gap of unknown length	
* 30184 44380: contig of 14157 bp in length	
* 44381 44480: gap of unknown length	
* 44481 69452: contig of 24972 bp in length	
* 69453 69552: gap of unknown length	
* 69553 102261: contig of 32709 bp in length	
* 102262 102361: gap of unknown length	
* 102362 105541: contig of 3180 bp in length	
* 105542 105641: gap of unknown length	
* 105642 129241: contig of 23600 bp in length	
* 129242 129341: gap of unknown length	
* 129342 133048: contig of 3707 bp in length	
* 133049 133148: gap of unknown length	
* 133149 150305: contig of 17156 bp in length	
* 150305 150404: gap of unknown length	
* 150405 153417: contig of 3013 bp in length	
* 153418 153517: gap of unknown length	
* 153518 180048: contig of 26531 bp in length.	
Location/Qualifiers	
1..180048	
/organism="Oryctolagus cuniculus"	
/mol_type="genomic DNA"	
/db_xref="taxon:9986"	
/clone="LB1-208620"	
/clone_lbb="LB1"	
/note="BAC resource: http://bacpac.chori.org/breed/New_Zealand_White "	
misc_feature	
1..20929	
/note="assembly_fragment"	
/clone_end:SP6	
vector_side:left"	
gap	
20930..21029	
/estimated_length=unknown	
misc_feature	
21030..30083	
/note="assembly_fragment"	
gap	
30084..30183	
/estimated_length=unknown	
misc_feature	
30184..44380	
/note="assembly_fragment"	
gap	
44381..44480	
/estimated_length=unknown	
misc_feature	
44481..69452	
/note="assembly_fragment"	
gap	
69453..69552	
/estimated_length=unknown	
misc_feature	
69553..102261	
/note="assembly_fragment"	
gap	
102262..102361	
/estimated_length=unknown	
misc_feature	
102362..105541	
/note="assembly_fragment"	
gap	
105542..105641	
/estimated_length=unknown	
misc_feature	
105642..129241	
/note="assembly_fragment"	
gap	
129242..129341	
/estimated_length=unknown	
misc_feature	
129342..133048	
/note="assembly_fragment"	
gap	
133049..133148	
/estimated_length=unknown	
misc_feature	
133149..150304	
/note="assembly_fragment"	

Sheety, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tinney, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villanar, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederstam, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 204401)
Worley, K.C.

Direct Submission
Submitted (09-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 204401)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23664495.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separate by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GNCM
Center clone name: CH230-212C17

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 182869 bases at least Q40
Consensus quality: 186192 bases at least Q30
Consensus quality: 188318 bases at least Q20
Estimated insert size: 187109; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved

1 203085: contig of 203085 bp in length
* 203086 203185: gap of unknown length
* 203186 204401: contig of 1216 bp in length.

Location/Qualifiers

1. 204401
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-212C17"

FEATURES
source

```

misc_feature      1..1221
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature      6359..7088
                  /note="clone_boundary
                  clone_end:T7
                  site:
                  end_sequence:BN352057"
                  29208..93076
                  /note="clone_boundary
                  clone_end:Sp6
                  site:
                  end_sequence:BN352058"
                  203086..203185
                  /estimated_length=unknown

ORIGIN
Query Match      81.6%; Score 20.4; DB 14; Length 204401;
Best Local Similarity 95.5%; Pred. No. 5.Se+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; gaps 0;

Qy 1 AAAAAAAAAATCCCAATTGAG 22
    |||||
Db 17805 AAAAAAAAAATCCCAATTGAG 17784

RESULT 38
AC151152 205446 bp DNA linear HTG 01-JUL-2005
LOCUS Bos taurus clone CH240-386E17, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
AC151152 GI:68265386
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Bos taurus (cow)
SOURCE Bos taurus
ORGANISM Bos taurus
REFERENCE
AUTHORS Muzny,D,Marie,M, Metzker,M, Lee,A, Abramson,S, Adams,C, Alder,J,
        Allen,C, Allen,H, Aisbrooks,S, Amin,A, Anguiano,D,
        Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,B, Baden,H,
        Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
        Biswal,K, Blair,J, Blankenship,K, Blych,P, Brown,M,
        Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
        Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,
        Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
        Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
        Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
        Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,
        Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
        Egan,A, Escotto,M, Eugene,C, Evans,C, Falls,T, Fan,G,
        Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
        Fraser,C,M, Gabriel,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
        Gebreyoung,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,M,
        Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
        Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
        Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogue,M,
        Hollins,B, Howells,S, Hulyk,S, Hume,J, Idelbird,D, Jackson,A,
        Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolyvet,A,
        Karpthy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
        Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
        Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
        Lorenznewe,L, Loulseg,H, Lozada,R,J, Lu,X, Ma,J,
        Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A,
        Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E,
        Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenan,E,
        Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
        Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L,
        Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Parks,K,
        Paoolelemon,O, Okunonu,G, Olarnpunsagoon,A, Pal,S, Parks,K,
        Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkoch,C,

```

COMMENT

```

TITLE JOURNAL
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-2004) Human Genome Sequencing Center, Department
        of Molecular and Human Genetics, Baylor College of Medicine, One
        Baylor Plaza, Houston, TX 77030, USA
        3 (bases 1 to 205446)
        Cow Genome Sequencing Consortium.
        Direct Submission
        Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
        of Molecular and Human Genetics, Baylor College of Medicine, One
        Baylor Plaza, Houston, TX 77030, USA
        On Jun 28, 2005 this sequence version replaced gi:52782475.
        The sequence in this assembly is a combination of BAC based reads
        and whole genome shotgun sequencing reads assembled using Atlas
        (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
        in the feature table below represents a scaffold in the Atlas
        assembly (a 'contig-scaffold'). Within each contig-scaffold,
        individual sequence contigs are ordered and oriented, and separated
        by fixed gaps filled with Ns to the estimated size. The sequence
        may extend beyond the ends of the clone and there may be sequence
        contigs within a contig-scaffold that consist entirely of whole
        genome shotgun sequence reads. Both end sequences and whole genome
        shotgun sequence only contigs will be indicated in the feature
        table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center Project name: FBSQ
Center clone name: CH240-386E17
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 199608 bases at least Q40
Consensus quality: 201216 bases at least Q30
Consensus quality: 202535 bases at least Q20
Estimated insert size: 205574; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2564: contig of 2564 bp in length
* 2565 3026: gap of 461 bp
* 3026 6558: contig of 3533 bp in length
* 6559 6817: gap of 259 bp

```

```

* 6818 12597: contig of 5780 bp in length
* 12598 12647: gap of 50 bp
* 12648 46697: contig of 34050 bp in length
* 46698 46747: gap of 50 bp
* 46748 79127: contig of 33380 bp in length
* 79128 79177: gap of 50 bp
* 79178 133293: contig of 54116 bp in length
* 133294 133343: gap of 50 bp
* 133344 192045: contig of 58702 bp in length
* 192046 192445: gap of unknown length
* 192446 195469: contig of 3324 bp in length
* 195470 195519: gap of 50 bp
* 195520 199196: contig of 3677 bp in length
* 199197 199572: gap of 376 bp
* 199573 201059: contig of 1487 bp in length
* 201060 201159: gap of unknown length
* 201160 202566: contig of 1407 bp in length
* 202567 202667: gap of unknown length
* 202667 204170: contig of 1504 bp in length
* 204171 204270: gap of unknown length
* 204271 205446: contig of 1176 bp in length.

```

```

FEATURES
Source
Location/Qualifiers
1..205446
/organism="Bos taurus"
/mol_type="Genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-386E17"
2565..3025
/estimated_length=461
6559..6817
/estimated_length=259
12598..12647
/estimated_length=50
46698..46747
/estimated_length=50
79128..79177
/estimated_length=50
133294..133343
/estimated_length=50
192046..192145
/estimated_length=50
195470..195519
/estimated_length=unknown
199197..199572
/estimated_length=50
201060..201159
/estimated_length=376
202567..202666
/estimated_length=unknown
204171..204270
/estimated_length=unknown

```

ORIGIN

```

Query Match      81.6%  Score 20.4;  DB 14;  Length 205446;
Best Local Similarity 95.5%;  Pred. NO. 5.5e+02;
Matches 21;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAG 22
Db 169734 AAAAAAAAAATCCCAATTCAG 169755

```

```

RESULT 39
AC128955
LOCUS
DEFINITION
AC128955 210179 bp DNA linear HTG 15-NOV-2002
Rattus norvegicus clone CH230-503A18, WORKING DRAFT SEQUENCE, 4
unordered pieces.
AC128955
AC128955.3 GI:25007477
HTG; HTG_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE

AUTHORS

COMMENT

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognath; Muridae; Muridae; Murinae; Rattus.

1 (bases 1 to 210179)

Mzyny,D,Marle, Metzker,M,Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Albrooke,S, Amin,A, Anguiano,D, Anyalabechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benhammed,F, Biewalot,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Cessar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Devila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedrich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Fall,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Giller,R, Grady,M, Guerra,M, Guevara,W, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,R, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M, Hollins,B, Howells,S, Huylk,S, Hume,J, Idlerbird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Ledow,H, Leyan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensbuewa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,B, Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwachileme,O, Okunonu,G, Olarunnsagoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L, L, Piazzi,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Saverly,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartberg,A, Sisson,I, Sitter,C,D, Smagis,D, Sneed,A, Sodergren,E, Song,X,Z, Sotelle,R, Sosa,J, Steidle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,J, Wang,O, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willison,R, Wlezyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R,A.

Direct Submission

Unpublished

2 (bases 1 to 210179)

Worley,K,C.

Direct Submission

Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 210179)

Direct Submission

Submitted (15-NOV-2002) Human Genome Sequencing Consortium.

Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23196185.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KBR6
Center clone name: CH230-503A18

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 16845 bases at least Q40
Consensus quality: 170523 bases at least Q30
Consensus quality: 171720 bases at least Q20
Estimated insert size: 169120; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a "working draft" sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 6701: contig of 6701 bp in length
* 6702 6801: gap of unknown length
* 6802 16784: contig of 9983 bp in length
* 16785 16884: gap of unknown length
* 16885 150992: contig of 134108 bp in length
* 150993 210179: gap of unknown length
* 151093 210179: contig of 59087 bp in length.
Location/Qualifiers

FEATURES
source 1. 210179
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-503A18"

gap 6702. 6801
/estimated_length=unknown

misc_feature 6802. 7901
/note="wgs_contig"

gap 16785. 16884
/estimated_length=unknown

misc_feature 16885. 18024
/note="wgs_contig"

gap 150993. 151092
/estimated_length=unknown

ORIGIN

Query Match 81.6%; Score 20.4; DB 14; Length 210179;
Best Local Similarity 95.5%; Pred. No. 5.5e+02;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTGACA 23
|||||
|||||

Db 142453 AAAAAAAAAAGTCAATTGACA 142474
|||||
|||||

RESULT 40
AC096528 216422 bp DNA linear HTG 10-MAY-2003
LOCUS AC096528
DEFINITION Rattus norvegicus clone CH230-156P9, WORKING DRAFT SEQUENCE.
ACCESSION AC096528
VERSION AC096528.5 GI:30522873
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus
Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 216422)

AUTHORS

Muzny,D,Marie, Metzker,M, Lee, Abtaramon, S, Adams, C, Alder, J, Allen, C, Allen, H, Alsbrooks, S, Amin, A, Anguiano, D, Anyalebechi, V, Aoyagi, A, Ayodeji, M, Baca, E, Baden, R, Baldwin, D, Bandaranaike, D, Barber, M, Barnstead, M, Benhmed, F, Biswal, K, Blair, J, Blankenburg, K, Blyth, P, Brown, M, Bryant, N, Buhay, C, Burch, P, Burrell, K, Calderon, E, Cardenas, V, Carter, K, Cavazos, I, Casar, H, Center, J, Chacko, J, Chavez, D, Chen, G, Chen, R, Chen, Y, Chen, Z, Chu, J, Cleveland, C, Cockrell, R, Cox, C, Coyle, M, Crease, A, D'Souza, L, Davila, M, L, Davis, C, Davy-Carroll, L, De Anda, C, Decker, D, Delgado, O, Denson, S, Deramo, C, Ding, Y, Din, H, Ditya, K, Draper, H, Dugan-Rocha, S, Dunn, A, Durbin, K, Duval, B, Eaves, K, Egan, A, Escotto, M, Eugene, C, Evans, C, A, Falls, T, Fan, G, Fernandez, S, Finley, M, Flagg, N, Forbes, L, Foster, M, Foster, P, Frazer, C, M, Gabisi, A, Ganta, R, Garcia, A, Garner, T, Garza, M, Gebregiorgis, E, Geer, K, Gill, R, Grady, M, Guerra, M, Guevara, W, Gunaratne, P, Haaland, W, Hamill, C, Hamilton, C, Hamilton, K, Harvey, Y, Havlak, P, Hawes, A, Henderson, N, Hernandez, J, Hernandez, R, Hines, S, Hladun, S, L, Hodgson, A, Hogue, M, Hollins, B, Howells, S, Huily, S, Hume, J, Idlebird, D, Jackson, A, Jackson, L, Jacob, L, Jiang, H, Johnson, B, Johnson, R, Jolivet, A, Karpathy, S, Kelly, S, Kelly, S, Khan, Z, King, L, Kovar, C, Kowis, C, Kraft, C, U, Lebow, H, Levay, J, Lewis, L, Li, Z, Liu, J, Liu, J, Liu, W, Liu, Y, London, P, Longacre, S, Lopez, J, Lorenzini, L, Louie, H, Lozano, R, J, Lu, X, Ma, J, Maheshwari, M, Mahindaratne, M, Mahmoud, M, Malloy, K, Mangum, A, Mangum, B, Mapa, P, Martin, K, Martin, R, Martinez, E, Mawhinney, S, McLeod, M, P, McNeill, T, Z, Meenan, E, Milosavljevic, A, Miner, G, Minja, E, Montemayor, J, Moore, S, Morgan, M, Morris, K, Morris, S, Munidasa, M, Murphy, M, Nair, L, Nankervis, C, Neal, D, Newton, N, Nguyen, N, Norris, S, Nwokedi, O, Okunolu, G, Olarinmoye, A, Pal, S, Parks, K, Pasternak, S, Paul, H, Perez, A, Perez, L, Plankoch, C, Plopper, F, Polidexter, A, Popovic, D, Primus, E, P, L, L, Puazo, M, Quiroz, J, Rachlin, E, Reeves, K, Regier, M, A, Reigh, R, Reilly, B, Reilly, M, Ren, Y, Reuter, M, Richards, S, Riggs, F, Rivers, C, Rodkey, T, Rojas, A, Rose, M, Rose, R, Ruiz, S, J, Sanders, W, Savary, G, Scherer, S, Scott, G, Shultman, S, Shen, H, Shetty, S, Shvartsbeyn, A, Sisson, I, Sitter, C, D, Sma, S, D, Sneed, A, Sodergren, R, Song, X, Z, Sorelle, R, Soes, J, Steinle, M, Strong, R, Sutton, A, Svatek, A, Tabor, P, Taylor, C, Taylor, T, Thomas, N, Thomas, S, Tingey, A, Trejos, Z, Umani, K, Valas, R, Vera, V, Villaseana, D, Waldron, L, Walker, B, Wang, J, Wang, Q, Wang, S, Warren, J, Warren, R, Wei, X, White, F, Williams, G, Willson, R, Wlezyk, R, Wooden, H, Worley, K, Wright, D, Wright, R, Wu, J, Yakub, S, Yen, J, Yoon, L, Yoon, V, Yu, F, Zhang, J, Zhou, J, Zhou, X, Zhou, S, Dunn, D, von Niederhausen, A, Weiss, R, Smith, D, R, Holt, R, A, Smith, H, O, Weinstein, G, and Gibbs, R, A.

TITLE

Unpublished

REFERENCE

2 (bases 1 to 216422)

AUTHORS

Worley, K, C.

TITLE

Direct Submission

JOURNAL

Submitted (18-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 216422)

AUTHORS

Rat Genome Sequencing Consortium.

TITLE

Direct Submission

JOURNAL

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 10, 2003 this sequence version replaced gi:23321936. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Aris

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GGCIC

Center clone name: CH230-156P9

Summary Statistics

Assembly program: Atlas 3.0

Consensus quality: 205458 bases at least Q40

Consensus quality: 206915 bases at least Q30

Estimated insert size: 220644; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* Location/Qualifiers
 1 216422: contig of 216422 bp in length.

FEATURES

SOURCE

1. 216422

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-156P9"

1. 11564

/note="wgs end extension"

clone_end:T7

/note="clone_boundary"

clone_end:T7

site:ECORI

end_sequence:BH259458"

complement(216189..216384)

/note="clone boundary"

clone_end:Sp6

site:ECORI

end_sequence:BH259459"

ORIGIN

Query Match 81.6%; Score 20.4; DB 14; Length 216422;

Best Local Similarity 95.5%; Pred. No. 5.4e+02;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAG 22

Db 99452 AAAAAAAAAAGTCCCAATTCAG 99473

RESULT 41

AC123017

LOCUS 220669 bp DNA linear HTG 15-NOV-2002

DEFINITION Rattus norvegicus clone CH230-282D11, *** SEQUENCING IN PROGRESS

ACCESSION

AC123017

VERSION

AC123017.4 GI:25007853

HTG: HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

REFERENCE

AUTHORS

Allen, C., Allen, H., Albrooke, S., Amin, A., Arguiano, D., Ayala-Becchi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bismail, N., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabris, A., Ganta, R., Garcia, A., Garner, T., Guevara, W., George, G., Geier, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, B., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulky, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Ledow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensbuewa, L., Louisedge, H., Lozada, R.J., Lu, X., Ma, J., Meshawari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., McInerney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minya, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwunonu, G., Olarinmoye, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Plambeck, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, T., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Soosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., and Gibbs, R.A.

TITLE

Unpublished

2 (bases 1 to 220669)

Worley, K.C.

Direct Submission

TITLE

Submitted (26-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 220669)

Rat Genome Sequencing Consortium.

Direct Submission

TITLE

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23908050.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZJY
Center clone name: CH230-282D11

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 205121 bases at least Q40
Consensus quality: 206790 bases at least Q30
Consensus quality: 207796 bases at least Q20
Estimated insert size: 213276; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter..
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 220669: contig of 220669 bp in length.
Location/Qualifiers

FEATURES

source

1. 220669
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-282D11"

misc_feature

1. 1865
/note="wgs end extension
clone_end:5p6"

misc_feature

/note="clone_boundary
clone_end:5p6
site:
end sequence:BZ238335"
complement(216546..217402)
/note="clone_boundary
clone_end:T7
site:
end sequence:BZ238333"
218299..220669
/note="wgs end extension
clone_end:T7"

misc_feature

complement(216546..217402)
/note="clone_boundary
clone_end:T7
site:
end sequence:BZ238333"
218299..220669
/note="wgs end extension
clone_end:T7"

misc_feature

complement(216546..217402)
/note="clone_boundary
clone_end:T7
site:
end sequence:BZ238333"
218299..220669
/note="wgs end extension
clone_end:T7"

ORIGIN

Query Match 81.6%; Score 20.4; DB 14; Length 220669;
Best Local Similarity 95.5%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAG 22

DB 180426 AAAAAAAAAAGTCCCAATTGAG 180447

RESULT 42

AC162948/c
LOCUS
DEFINITION
MUS musculus BAC clone RP23-14N2 from chromosome 9, complete
sequence.
AC162948
AC162948.4 GI:71533526
HTG.
MUS musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
1 (bases 1 to 221791)
Schreiber, B., Lewis, S., Levy, A. and Haglund, K.
The sequence of Mus musculus BAC clone RP23-14N2
Unpublished (2001)
2 (bases 1 to 221791)
Wilson, R. K.
Direct Submisson
Submitted (02-JUN-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 221791)
Wilson, R. K.
Direct Submisson
Submitted (24-JUN-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 221791)
Wilson, R. K.
Direct Submisson
Submitted (30-JUN-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 30, 2005 this sequence version replaced gi:68163763.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_BA0014N02

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e. phred quality
>30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone, fosmid clone or direct clone walk sequence.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to obtain the consensus sequence. The
assembly was confirmed by restriction digest.
This finishing standard has slightly changed from the previous
Human standard. Specifically, standards for regions of low sequence
complexity (such as dinucleotide repeats and small unit tandem
repeats) have been relaxed. These regions are very prevalent in the
mouse genome, and the return on extended finishing efforts is
minimal.

If a sequence meets the criteria of the above statement, it needs
no comments or tags. If the criteria are not met, such as ambiguous
bases, then the region is duly annotated.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:

The BAC library has been constructed by Kazutoyo Osegawa and
Minako Tateo in the laboratory of Pletier de Jong
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or

brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.reagen.com>) or Pletter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

Location/Qualifiers

1..221791
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="9"
/clone="RP23-14N2"
/clone_1fb="RPCT-23"
10832..10997
/note="Sequence derived from one plasmid subclone."

ORIGIN

Query Match 81.6% Score 20.4; DB 9; Length 221791;
Best Local Similarity 95.5%; Pred. No. 5.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAG 22
Db 117446 AAAAAAAAAAGTCCCAATTCAG 117425

AC096159 237549 bp DNA linear HTG 10-MAY-2003
Rattus norvegicus clone CH230-11D18, *** SEQUENCING IN PROGRESS
*** 3 unordered pieces.

AC096159.7 GI:30522589
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 237549)

REFERENCE
AUTHORS
Allen, C., Allen, H., Metzger, M., Lee, A., Adams, C., Alder, J.,
Muzny, D., Marie, E., Alsebrook, S., Amin, A., Anguiano, D.,
Ayala, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaru, A., Barker, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blythe, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devita, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, C., Dierker, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabriel, A., Gante, R., Garcia, A., Garner, T., Garza, M.,
Gehrke, G., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,
Henderson, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
Hollins, B., Howell, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, D., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, D.,
Lorenshwar, L., Loulsegod, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallow, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,
Millojavlevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Ngundasa, M., Norris, S.,
Nwachukwu, O., Okunolu, G., Olarnunsgoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Piazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Savary, G., Scherer, S., Scott, C., Shalman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smales, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Swatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villars, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 237549)
Worley, K.C.

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237549)
Baylor Genome Sequencing Consortium.

Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:24635746.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GEOA
Center clone name: CH230-11D18
----- Summary Statistics
Assembly program: Atlas 3.0
Consensus quality: 224484 bases at least Q40
Consensus quality: 227184 bases at least Q30
Consensus quality: 229173 bases at least Q20
Estimated insert size: 240500; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 3 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 234524: contig of 234524 bp in length
* 234525 234624: gap of unknown length
* 234625 235873: contig of 1249 bp in length
* 235874 235973: gap of unknown length

```

FEATURES      * 235974 237549: contig of 1576 bp in length.
               Location/Qualifiers
                source
                  1. 237549
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-11D18"
                     433..1087
                     /note="clone boundary"
                     clone_end:T7
                     site:ECORI
                     end_sequence:BH340049"
                     32014..33725
                     /note="wgs contig"
                     60142..62249
                     /note="clone boundary"
                     clone_end:Sp6
                     site:ECORI
                     end_sequence:BH340050"
                     215407..217527
                     /note="wgs end extension"
                     clone_end:Sp6"
                     234525..234624
                     /estimated_length=unknown
                     235874..235973
                     /estimated_length=unknown

ORIGIN
Query Match      81.6%; Score 20.4; DB 14; Length 237549;
Best Local Similarity 95.5%; Pred. No. 5,3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  AAAAAAAAACTCCCAATTCAGA 23
        |||||
        2192 AAAAAAAAAAGTCAATTGAGA 2171

RESULT 44
AC132157      239661 bp      DNA      linear      HTG 19-NOV-2002
LOCUS      AC132157
DEFINITION      Rattus norvegicus clone CH230-96C17, *** SEQUENCING IN PROGRESS
ACCESSION      AC132157
VERSION      AC132157.4 GI:25073585
KEYWORDS      HTG; HTGS_PHASeI; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
               Rukavota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
               Sciurognathi; Muridae; Muridae; Murinae; Rattus.
               1 (bases 1 to 239661)
               Allen, C., Allen, H., Albrooke, S., Amin, A., Anguiano, D.,
               Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, R., Baden, H.,
               Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
               Biewick, K., Blair, J., Blankenburg, K., Blych, P., Brown, M.,
               Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
               Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
               Chacco, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
               Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
               Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
               Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
               Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
               Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
               Fernandez, S., Finley, M., Flagge, N., Forbes, L., Foster, M., Foster, P.,
               Frazer, C.M., Gabies, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
               Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
               Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, J.,
               Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
               Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
               Hollins, B., Howells, S., Hu, Y., Hume, J., Idelberg, D., Jackson, A.,
               Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
               Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

```

```

TITLE      JOURNAL
REFERENCE   2 (bases 1 to 239661)
AUTHORS     Rat Genome Sequencing Consortium.
TITLES      Direct Submision
JOURNALS    Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 239661)
            Rat Genome Sequencing Consortium.
            Direct Submision
            Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Nov 19, 2002 this sequence version replaced gi:22855857.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Actas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Actas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by extend gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.

            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: KBAC
            Center clone name: CH230-96C17
            ----- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 224069 bases at least Q40
            Consensus quality: 226850 bases at least Q30
            Consensus quality: 228630 bases at least Q20
            Estimated insert size: 227594; sum-of-contigs estimation
            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 229796: contig of 229796 bp in length
* 229797 229896: gap of unknown length
* 229897 235391: contig of 5495 bp in length
* 235392 235491: gap of unknown length
* 235492 237060: contig of 1569 bp in length
* 237061 237160: gap of unknown length
* 237161 238259: contig of 1099 bp in length
* 238260 238359: gap of unknown length
* 238360 239661: contig of 1302 bp in length.
*
* Location/Qualifiers
*   1..239661
*     /organism="Rattus norvegicus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:10116"
*     /clone="CH230-96C17"
*     1..2251
*       /note="wgs contig"
*       229797..229896
*         /estimated_length=unknown
*         229897..231549
*           /note="wgs contig"
*           235392..235491
*             /estimated_length=unknown
*             237061..237160
*               /estimated_length=unknown
*               238260..238359
*                 /estimated_length=unknown

```

ORIGIN

Query Match

Best Local Similarity 81.6%; Score 20.4; DB 14; Length 239661;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 AAAAAAAAAAGTCCCATTCAG 22

Db 5499 AAAAAAAAAAGTCCCATTCAG 5520

RESULT 45

AC111918/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-32A6, *** SEQUENCING IN PROGRESS ***

AC111918 245581 bp DNA linear HTG 09-NOV-2002

2 unordered pieces.

AC111918

AC111918.4 GI:24818638

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Rattus.

1 (bases 1 to 245581)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blych, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Cesari, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Detam, C., Ding, Y., Dinh, H., Diya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, J., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huliy, S., Hume, J., Idelbrad, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensunewa, L., Louisedge, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McKell, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweme, O., Okwodu, G., Olarnunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Saverly, G., Scherer, S., Scott, C., Shatsman, S., Shen, H., Shetty, D., Shvartsbeyn, A., Sisson, T., Sitter, C. D., Smaj, D., Sneed, A., Sodergren, E., Song, X., Z., Sorrelle, R., Soze, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 245581)

Worley, K. C.

Direct Submission

Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 245581)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:23265279.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GONN

Center clone name: CH230-32A6

----- Summary Statistics

Assembly program: Phrap version 0.990329

Consensus quality: 233622 bases at least Q40

Consensus quality: 236336 bases at least Q30
Consensus quality: 238280 bases at least Q20
Estimated insert size: 242327; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 243199: contig of 243199 bp in length
* 243200 243299: gap of unknown length
* 243300 245581: contig of 2282 bp in length.
Location/Qualifiers
1. 245581
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-32A6"
1. 1597
/note="wgs_end_extension
clone_end:Sp6"
2721. 3077
/note="clone boundary
clone_end:Sp6
site:ECORI
end_sequence:BH325890"
136374. 216919
/note="clone boundary
clone_end:T7
site:ECORI
end_sequence:BH325889"
241852. 243199
/note="wgs_end_extension
clone_end:T7"
243200. 243299
/estimated_length=unknown

ORIGIN
Query Match 81.6%; Score 20.4; DB 14; Length 245581;
Best Local Similarity 95.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGA 23
|||||
DB 203304 AAAAAAAAAAGTCCCAATTCAGA 203283

RESULT 46
AC164434/c 255682 bp DNA linear HTG 29-JUN-2005
LOCUS AC164434
DEFINITION Mus musculus chromosome 9 clone RP23-341E21, WORKING DRAFT
ACCESSION AC164434
VERSION AC164434.2 GI:68303797
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 255682)
Wilson,R.K.
REFERENCE 1 The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 255682)
Wilson,R.K.
TITLE Direct Submission

JOURNAL Submitted (21-JUN-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 255682)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jun 29, 2005 this sequence version replaced gi:68051922.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0311E21
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 249145 bases at least Q40
Consensus quality: 250855 bases at least Q30
Consensus quality: 252196 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1331: contig of 1331 bp in length
* 1332 1431: gap of unknown length
* 1432 2520: contig of 1089 bp in length
* 2521 2620: gap of unknown length
* 2621 3968: contig of 1348 bp in length
* 3969 4068: gap of unknown length
* 4069 5338: gap of 1270 bp in length
* 5339 5438: gap of unknown length
* 5439 6745: contig of 1307 bp in length
* 6746 6845: gap of unknown length
* 6846 7922: contig of 1077 bp in length
* 7923 8022: gap of unknown length
* 8023 10043: contig of 2021 bp in length
* 10044 10143: gap of unknown length
* 10144 12802: contig of 2659 bp in length
* 12803 12902: gap of unknown length
* 12903 14708: contig of 1806 bp in length
* 14709 14808: gap of unknown length
* 14809 14982: gap of 5174 bp in length
* 14983 19983: gap of unknown length
* 19984 20082: gap of unknown length
* 20083 30552: contig of 10470 bp in length
* 30553 30652: gap of unknown length
* 30653 49707: contig of 19055 bp in length
* 49708 49807: gap of unknown length
* 49808 92962: contig of 43155 bp in length
* 92963 93062: gap of unknown length
* 93063 133392: contig of 40330 bp in length
* 133393 133492: gap of unknown length
* 133493 255682: contig of 122190 bp in length.
Location/Qualifiers
1. 255682
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="9"
/clone="RP23-341E21"
1. 1331
/note="assembly_name:Contig29"

misc_feature
1. 1331

```

gap      1332..1431
         /estimated_length=unknown
misc_feature 1432..2520
         /note="assembly_name:Contig42"
gap      2521..2620
         /estimated_length=unknown
misc_feature 2621..3968
         /note="assembly_name:Contig43"
gap      3969..4068
         /estimated_length=unknown
misc_feature 4069..5338
         /note="assembly_name:Contig44"
gap      5339..5438
         /estimated_length=unknown
misc_feature 5439..6745
         /note="assembly_name:Contig60"
gap      6746..6845
         /estimated_length=unknown
misc_feature 6846..7922
         /note="assembly_name:Contig63"
gap      7923..8022
         /estimated_length=unknown
misc_feature 8023..10043
         /note="assembly_name:Contig64"
gap      10044..10143
         /estimated_length=unknown
misc_feature 10144..12802
         /note="assembly_name:Contig65"
gap      12803..12902
         /estimated_length=unknown
misc_feature 12903..14708
         /note="assembly_name:Contig66"
gap      14709..14808
         /estimated_length=unknown
misc_feature 14809..19982
         /note="assembly_name:Contig67"
gap      19983..20082
         /estimated_length=unknown
misc_feature 20083..30552
         /note="assembly_name:Contig68"
gap      30553..30652
         /estimated_length=unknown
misc_feature 30653..49707
         /note="assembly_name:Contig69"
gap      49708..49807
         /estimated_length=unknown
misc_feature 49808..92962
         /note="assembly_name:Contig70"
gap      92963..93062
         /estimated_length=unknown
misc_feature 93063..133392
         /note="assembly_name:Contig71"
gap      133393..133492
         /estimated_length=unknown
misc_feature 133493..255682
         /note="assembly_name:Contig72"
ORIGIN
Query Match      81.6%; Score 20.4; DB 14; Length 255682;
Best Local Similarity 95.5%; Pred. No. 5.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 AAAAAAAAAAGTCCCAATTCAG 22
         |||||
Db      220430 AAAAAAAAAAGTCCCAATTCAG 220409

```

```

RESULT 47
BV175580/c      201 bp      DNA      linear      STS 10-JUN-2004
DEFINITION      segm81517 Human DNA (sequenom) Homo sapiens STS genomic, sequence
tagged site.
ACCESSION      BV175580

```

```

VERSION      BV175580.1 GI:48011203
KEYWORDS      STS.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE
AUTHORS      1 (bases 1 to 201)
              Nelson,R.M., Marnettloef,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
              Cantor,C.R. and Braun,A.
TITLE      Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
Regions
JOURNAL      Genome Res. (2004) In press
COMMENT
Contact: Andreas Braun
Pharmaceuticals division
Sequenom, Inc.
3595 John Hopkins Court, San Diego, CA 92121, USA
Tel: 18582029018
Fax: 18582029020
Email: abraun@sequenom.com
Primer A: No primer sequence submitted
Primer B: No primer sequence submitted
STS size: 201.
FEATURES
source
1..201
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="Human DNA (Sequenom)"
<1..>201
ORIGIN
STS
Query Match      80.8%; Score 20.2; DB 10; Length 201;
Best Local Similarity 88.0%; Pred. No. 5.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      1 AAAAAAAAAAGTCCCAATTCAGATA 25
         |||||
Db      165 AAAAAAAAAAGTACATTCAGATA 141

```

```

RESULT 48
BV442141/c      422 bp      DNA      linear      STS 01-FEB-2005
LOCUS      S2376151F89.T0 PortugueseWaterDog Canis familiaris STS genomic,
DEFINITION      sequence tagged site.
BV442141
ACCESSION      BV442141.1 GI:57860149
VERSION
SOURCE
KEYWORDS      Canis familiaris (dog)
ORGANISM      Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
REFERENCE
AUTHORS      1 (bases 1 to 422)
              Lindblad-Toh,K.
TITLE      The genome sequence of Canis familiaris
JOURNAL      Unpublished (2004)
COMMENT
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersl@genome.wi.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 422
Protocol:
WGS-discovery (WGS):
Paired-end low-coverage whole genome shotgun reads were generated

```

from 9 breeds (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador Retriever, English Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese Water Dog -100,000 each) and five other canids (Chinese, Alaskan, Indian and Spanish Gray Wolf as well as the Californian Coyote).

The WGS reads were placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 863872 reads were annotated as STSs and 485941 SNPs were annotated with alleles from the boxer and the breed or canid from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

WGA-discovery (WGA) of Boxer/Poodle SNPs:

A second set of SNPs was generated using a similar methodology except that the contigs from the 1.5x poodle assembly (Kirkness 2003) were used instead of WGS reads. Since this sequence lacked base quality scores, arbitrary quality scores of phred 40 were assigned before the poodle sequence was placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated with alleles from the boxer and the poodle. The validation rate for these SNPs was estimated at approximately 98%.

Internal-WGA-discovery (I-WGA):

A third set of SNPs were discovered by comparing reads in the WGA assembly. SNPs were defined as mismatch positions that had a base quality of ≥ 30 on both reads in a region that aligned without gaps, and with at most one additional mismatch in the ten flanking bases. For each allele, at least one additional read had to confirm it. 731476 SNPs were annotated with alleles between the two boxer alleles. The validation rate for these SNPs was estimated at approximately 98%.

FEATURES

source

```
1..422
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="PortugueseWaterDog"
/db_xref="taxon:9615"
/map="15 22-371 5960211-5959862"
/clone_11b="PortugueseWaterDog"
<1..>579
```

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 422;
Best Local Similarity 88.0%; Pred. No. 4.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25

DB 411 AAAAAAAAAAGGCCCAATACCAATA 387

RESULT 49

BV340931

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BV340931 579 bp DNA linear STS 27-JAN-2005
S2306519F11.T0 Rottweiler Canis familiaris STS genomic, sequence
tagged site.
BV340931
BV340931.1 GI:57541234
STS.
Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;

REFERENCE

1 (bases 1 to 579)
Lindblad-Toh, K.
The genome sequence of Canis familiaris
Unpublished (2004)

COMMENT

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kers@genome.wi.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 579
Protocol:

WGS-discovery (WGS):
Paired-end low-coverage whole genome shotgun reads were generated from 9 breeds (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador Retriever, English Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese Water Dog -100,000 each) and five other canids (Chinese, Alaskan, Indian and Spanish Gray Wolf as well as the Californian Coyote).

The WGS reads were placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 863872 reads were annotated as STSs and 485941 SNPs were annotated with alleles from the boxer and the breed or canid from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

WGA-discovery (WGA) of Boxer/Poodle SNPs:

A second set of SNPs was generated using a similar methodology except that the contigs from the 1.5x poodle assembly (Kirkness 2003) were used instead of WGS reads. Since this sequence lacked base quality scores, arbitrary quality scores of phred 40 were assigned before the poodle sequence was placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated with alleles from the boxer and the poodle. The validation rate for these SNPs was estimated at approximately 98%.

Internal-WGA-discovery (I-WGA):

A third set of SNPs were discovered by comparing reads in the WGA assembly. SNPs were defined as mismatch positions that had a base quality of ≥ 30 on both reads in a region that aligned without gaps, and with at most one additional mismatch in the ten flanking bases. For each allele, at least one additional read had to confirm it. 731476 SNPs were annotated with alleles between the two boxer alleles. The validation rate for these SNPs was estimated at approximately 98%.

FEATURES

source

```
1..579
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Rottweiler"
/db_xref="taxon:9615"
/map="15 22-531 31581923-31582431"
/clone_11b="Rottweiler"
<1..>579
```

ORIGIN

STS

Query Match 80.8%; Score 20.2; DB 10; Length 579;
Best Local Similarity 88.0%; Pred. No. 3.7e+03;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 429 AAAAAAAAAATCCCAATTCAGAAA 453

RESULT 50

BV159860/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2003)

Contact: Jonathan Usuka

Roche Palo Alto Genetics and Genomics Department

Roche Palo Alto

3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA

Tel: 6508555807

Email: Jonathan.Usuka@roche.com

Primer A: No primer submitted with this STS

Primer B: No primer submitted with this STS.

Location/Qualifiers

1..582

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/map="13-20810-21386-AU590614.8.1.210845"

/clone_id="Roche Palo Alto"

/note="SNPs developed from assay sequences derived from 15

different strains of mice (as of October 1, 2003). Those

strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/Osnu,

BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,

MRL/MpJ, NZB/BinJ, NZW/LacJ, SPRET/Ei."

<1..>582

STS

ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 10; Length 582;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 494 AAAAAAAAAAGTCCCAATTCAGATA 470

Search completed: December 14, 2005, 11:11:32
Job time : 877.8 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 ; Search time 203.2 Seconds
(without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-7

Perfect score: 25

Sequence: 1 aaaaaaaaaagtcctcaatcagata 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

N_Geneseq 21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	21.8	87.2	5933	6 ABK39977	Abk39977 Human che
C 2	21.8	87.2	10151	6 ABK39433	Abk39433 Human imm
C 3	20.2	80.8	12542	4 AAS26800	Aas26800 Human gen
C 4	20.2	80.8	12542	8 ABX74149	Abx74149 Human nov
C 5	20.2	80.8	13558	4 AAI66274	Aai66274 Murine ap
C 6	20.2	80.8	15320	4 AAK78660	Aak78660 Human imm
C 7	20.2	80.8	15320	4 AAK78662	Aak78662 Human imm
C 8	20.2	80.8	17893	6 ABL33364	Abk33364 Human imm
C 9	20.2	80.8	109586	11 ACN43994	Acn43994 Human gen
C 10	19.8	79.2	348	6 ABL66419	Abk66419 Lung canc
C 11	19.8	79.2	588	12 ADL13116	Adl13116 Human sec
C 12	19.8	79.2	6974	11 ADN95749	Adn95749 Human bre
C 13	19.8	79.2	6974	12 ADO19267	Ado19267 Human pro
C 14	19.8	79.2	6974	12 ADO19264	Ado19264 Human pro
C 15	19.8	79.2	6976	13 ACN38387	Acn38387 Human sof
C 16	19.8	79.2	7032	8 ABX77603	Abx77603 Diferent
C 17	19.8	79.2	7835	12 ADO23503	Ado23503 Human sof
C 18	19.8	79.2	10514	4 AAL37105	Aal37105 Human mus
C 19	19.8	79.2	10514	8 ABX60093	Abx60093 cDNA enco

C 20	19.8	79.2	10514	12 ADJ30843	Adj30843 Human mus
C 21	19.4	77.6	991	13 ADT18127	Adt18127 Plant cDN
C 22	19.4	77.6	17131	3 AAZ60888	Aaz60888 DNA enco
C 23	19.4	77.6	51001	12 ADJ10262	Adj10262 Human ger
C 24	19.4	77.6	235033	2 AAV57926	Aav57926 Hereditar
C 25	19.4	77.6	237326	2 AAV57903	Aav57903 Hereditar
C 26	19.2	76.8	774	14 AEB67455	Aeb67455 Rice geno
C 27	19.2	76.8	870	4 AAK87145	Aak87145 Human imm
C 28	19.2	76.8	978	4 AAK87145	Aak87145 Human imm
C 29	19.2	76.8	1397	6 AAK87147	Aak87147 Human imm
C 30	19.2	76.8	1821	6 ABZ70778	Abz70778 Human par
C 31	19.2	76.8	2567	13 ADX55035	Adx55035 Plant ful
C 32	19.2	76.8	2690	12 ADQ64807	Adq64807 Novel hum
C 33	19.2	76.8	4411	12 ADH69515	Adh69515 Pig dopam
C 34	19.2	76.8	5097	6 AAL44712	Aal44712 Human glu
C 35	19.2	76.8	5503	3 AAC74570	Aac74570 Human ORF
C 36	19.2	76.8	5628	4 AAS36240	Aas36240 Human car
C 37	19.2	76.8	5628	4 AAL03224	Aal03224 Human rep
C 38	19.2	76.8	5628	10 ADE46934	Ade46934 Human car
C 39	19.2	76.8	5628	13 ADJ08352	Adj08352 Human car
C 40	19.2	76.8	6134	13 ABD32925	Abd32925 Human can
C 41	19.2	76.8	6138	13 ABD32926	Abd32926 Human can
C 42	19.2	76.8	31271	4 ABK26680	Abk26680 Drosophi
C 43	19.2	76.8	65454	12 ADN01773	Adn01773 Human hun
C 44	19.2	76.8	66315	12 ADQ97351	Adq97351 Human can
C 45	19.2	76.8	88759	13 ABD32923-7	Abd32923-7 of
C 46	19.2	76.0	67674	12 ADP81772	Adp81772 Human MD-
C 47	18.8	75.2	2971	12 ADP06628	Adp06628 Rice ale
C 48	18.8	75.2	3696	4 AAC90666	Aac90666 Stradect
C 49	18.8	75.2	5642	4 AAK69744	Aak69744 Human imm
C 50	18.8	75.2	5642	5 AAS34694	Aas34694 Human DNA
C 51	18.8	75.2	7421	12 ADM47692	Adm47692 Polynucle
C 52	18.8	75.2	28564	10 ADD48585	Add48585 Human gen
C 53	18.8	75.2	28564	10 ADD46617	Add46617 Human gen
C 54	18.8	75.2	34389	11 ACN44764	Acn44764 Mouse gen
C 55	18.6	74.4	406	5 AAF66619	Aaf66619 Novel hum
C 56	18.6	74.4	410	4 AAL87328	Aal87328 Human pol
C 57	18.6	74.4	431	5 ADL63458	Adl63458 Human ova
C 58	18.6	74.4	435	13 ADV40161	Adv40161 Rat cardl
C 59	18.6	74.4	441	5 ABV17722	Abv17722 Human pro
C 60	18.6	74.4	514	5 ADL4328	Adl4328 Human ova
C 61	18.6	74.4	717	4 AAH07799	Aah07799 Human cDN
C 62	18.6	74.4	730	10 ADE53745	Ade53745 Human pro
C 63	18.6	74.4	945	4 AAL20961	Aal20961 Human bre
C 64	18.6	74.4	1240	13 ADX64020	Adx64020 Plant ful
C 65	18.6	74.4	1315	7 ADZ74629	Adz74629 Arabidops
C 66	18.6	74.4	1484	14 ADY39424	Ady39424 Maize pla
C 67	18.6	74.4	1965	4 ABA45342	Ab45342 Human bre
C 68	18.6	74.4	1965	4 ABA45342	Ab45342 Human bre
C 69	18.6	74.4	1965	6 ABS04085	Abs04085 Human liv
C 70	18.6	74.4	2000	6 ABZ17097	Abz17097 Arabidops
C 71	18.6	74.4	2348	4 AAH14971	Aah14971 Human cDN
C 72	18.6	74.4	2546	4 AAK64601	Aak64601 Human imm
C 73	18.6	74.4	2644	6 ABQ73966	Abq73966 Human 3-t
C 74	18.6	74.4	2663	4 ABL25550	Abk25550 Drosophi
C 75	18.6	74.4	2802	4 AAK62817	Aak62817 Human imm
C 76	18.6	74.4	3150	4 AAS31543	Aas31543 Human DNA
C 77	18.6	74.4	3150	6 ABO66867	Abq66867 Human pol
C 78	18.6	74.4	3150	10 ADCL1154	Adcl1154 Human DNA
C 79	18.6	74.4	3195	4 ABL16002	Abk16002 Drosophi
C 80	18.6	74.4	3223	2 AAO62917	Aao62917 hrRNP U P
C 81	18.6	74.4	3410	4 AAK85167	Aak85167 Human imm
C 82	18.6	74.4	3497	8 ACC81115	Acc81115 Human and
C 83	18.6	74.4	3497	12 ADK14136	Adk14136 Human aut
C 84	18.6	74.4	3497	14 ADY14732	Ady14732 DNA enco
C 85	18.6	74.4	3497	14 ADZ49280	Adz49280 Insulin s
C 86	18.6	74.4	3561	6 ABA59840	Abn59840 Novel hum
C 87	18.6	74.4	3598	4 AAH14467	Aah14467 Human cDN
C 88	18.6	74.4	3598	13 ADZ25896	Adz25896 Breast ca
C 89	18.6	74.4	3759	4 ABL08968	Abk08968 Drosophi
C 90	18.6	74.4	3759	4 AAS57090	Aas57090 DNA enco
C 91	18.6	74.4	3759	10 ADC35776	Adc35776 Drosophi
C 92	18.6	74.4	4083	6 ABB53001	Abb53001 cDNA enco

C 93	18.6	74.4	4169	4	ABL16690	Ab116690 Drosophila
C 94	18.6	74.4	5562	6	ABL33934	Ab133934 Human imm
C 95	18.6	74.4	5562	6	ABL92294	Ab192294 Chemical
C 96	18.6	74.4	5562	6	ABL49363	Ab149363 Human pol
C 97	18.6	74.4	5933	6	ABK39978	Abk39978 Human che
C 98	18.6	74.4	6486	6	ABK67050	Abk67050 Human ang
C 99	18.6	74.4	6650	6	ABL32639	Ab132639 Human imm
C 100	18.6	74.4	8064	6	ABL33387	Ab133387 Human imm
C 101	18.6	74.4	9510	6	AA546438	AA546438 Tumour su
C 102	18.6	74.4	9510	6	ABL34563	Ab134563 Human met
C 103	18.6	74.4	9510	7	AD598824	Ad598824 Complemen
C 104	18.6	74.4	10151	6	ABL14432	Ab134432 Human imm
C 105	18.6	74.4	11147	8	AB210007	Ab210007 Haematopo
C 106	18.6	74.4	11147	8	AB210153	Ab210153 Haematopo
C 107	18.6	74.4	11169	6	ABK67095	Abk67095 Human ang
C 108	18.6	74.4	11422	6	ABK39936	Abk39936 Human che
C 109	18.6	74.4	11422	6	ABL32218	Ab132218 Human imm
C 110	18.6	74.4	12673	4	AAK59115	AAK59115 Human imm
C 111	18.6	74.4	12673	4	AAK78469	AAK78469 Human imm
C 112	18.6	74.4	14147	6	AA546743	AA546743 Tumour su
C 113	18.6	74.4	14147	6	ABK33955	Abk33955 Human DNA
C 114	18.6	74.4	14147	8	ADA20386	Ada20386 Prostate
C 115	18.6	74.4	14147	8	ADA84193	Ada84193 Human ren
C 116	18.6	74.4	14147	10	AD554235	Ad554235 Pretreat
C 117	18.6	74.4	14147	10	AD554087	Ad554087 Pretreat
C 118	18.6	74.4	14147	10	AD584087	Ad584087 Human lym
C 119	18.6	74.4	14147	10	AD584163	Ad584163 Human lym
C 120	18.6	74.4	14147	13	AD589249	Ad589249 Oligonuc
C 121	18.6	74.4	14147	13	AD589523	Ad589523 Oligonuc
C 122	18.6	74.4	14147	4	ABL28744	Ab128744 Drosophila
C 123	18.6	74.4	15500	4	AA546508	AA546508 Tumour su
C 124	18.6	74.4	16033	6	ABL33404	Ab133404 Human imm
C 125	18.6	74.4	18183	4	AA546280	AA546280 Tumour su
C 126	18.6	74.4	18183	6	ABK31159	Abk31159 Signal tr
C 127	18.6	74.4	18183	6	ABL70112	Ab170112 Chemical
C 128	18.6	74.4	19131	8	AA546716	AA546716 Tumour su
C 129	18.6	74.4	20142	8	ACA19757	ACA19757 Prokaryot
C 130	18.6	74.4	23815	4	AAK85169	AAK85169 Human imm
C 131	18.6	74.4	23815	4	AAK68678	AAK68678 Human imm
C 132	18.6	74.4	32185	11	ACN44604	ACN44604 Mouse gen
C 133	18.6	74.4	33805	11	ACN44836	ACN44836 Mouse gen
C 134	18.6	74.4	35641	6	ABL64428	Ab164428 Stomach c
C 135	18.6	74.4	35641	6	ABN95727	Abn95727 Gene #222
C 136	18.6	74.4	46181	14	AD213226	Ad213226 Murine ca
C 137	18.6	74.4	53321	14	AD259527	Ad259527 Secondary
C 138	18.6	74.4	61791	13	ABD33484	ABD33484 Murine ca
C 139	18.6	74.4	65524	13	ADV15291	Adv15291 Human oes
C 140	18.6	74.4	65524	13	ADV15297	Adv15297 Human oes
C 141	18.6	74.4	70019	13	ABD33601	ABD33601 Murine ca
C 142	18.6	74.4	76406	12	ADQ59203	Adq59203 MSI-H ca
C 143	18.6	74.4	77834	11	ACN44076	ACN44076 Mouse gen
C 144	18.6	74.4	95269	6	ABK67195_4	ABK67195_4 of
C 145	18.6	74.4	97835	6	ABK84796_	ABK84796 Human CDN
C 146	18.6	74.4	110000	2	AA20248_01	AA20248_01 of
C 147	18.6	74.4	110000	2	AA20248_02	AA20248_02 of
C 148	18.6	74.4	110000	6	ABQ69245_13	ABQ69245_13 of
C 149	18.6	74.4	110000	12	ADH69807_0	ADH69807_0 of
C 150	18.6	74.4	110000	14	AD245062_04	AD245062_04 of

ALIGNMENTS

RESULT 1

ABK39977/c
ID ABL34433 standard; DNA: 5933 BP.XX ABL34433;
XX AC ABL34433;
XX DT 21-MAY-2002 (first entry)
XX DE Human chemically pretreated gene sequence #30 strand 1.

KW	Human; de; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW	Cytosolic; ALDH6; CYP11A; CYP11B; DYPD; EPHX2; OCLN; TXNRD1;
KW	UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
OS	Homo sapiens.
PN	WO200202806-A2.
PD	10-JAN-2002.
PF	29-JUN-2001; 2001WO-BP007470.
PR	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
PA	(EPIC-) EPIGENOMICS AG.
PI	Olek A, Piepenbrock C, Berlin K;
DR	WPI; 2002-154757/20.
XX	New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT	useful for detecting cytosine methylation state of genes associated with
PT	pharmacogenomics and for therapy of diseases e.g. cancer.
PS	Claim 1; SEQ ID NO 59; 24pp; English.
XX	The invention relates to a nucleic acid comprising a sequence at least 18
CC	bases in length of a segment of the chemically pretreated DNA of genes
CC	associated with pharmacogenomics according to one of the sequences of the
CC	genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B (NM_000497), CYP3A3
CC	(NM_000776 and NM_017460), DYPD (NM_000110), EPHX2 (NM_004996),
CC	(NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996),
CC	(NM_019900), NM_019901, NM_019902, NM_019867, NM_019898, NM_019899) and
CC	their complementary sequences, or a sequence (St) chosen from 87
CC	sequences and their complements. The chemical pretreatment is bisulphite
CC	treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC	Also included are an oligomer (11) in particular an oligonucleotide or a
CC	peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC	base sequence having a length of 9 nucleotides which hybridises to or is
CC	identical to a chemically pretreated DNA of genes associated with
CC	pharmacogenomics and their complements, arranged in an array for
CC	analysing diseases associated with the methylation state (CpG) and/or
CC	detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC	oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC	and their complements is useful for diagnosis and therapy of solid
CC	tumours and cancer. The present sequence represents one the 87 DNA
CC	sequences or its complement. Note: The sequence data for this patent did
CC	not form part of the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
SQ	Sequence 5933 BP; 1686 A; 141 C; 1346 G; 2760 T; 0 U; 0 Other;
QY	Query Match 87.2%; Score 21.8; DB 6; Length 5933;
QY	Best Local Similarity 92.0%; P-adj. No. 49;
QY	Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DY	1 AAAAAAAAAAGTCCCAATTCAGATA 25
DY	2331 AAAAAAAAAAATCCCAATTCAGATA 2307

RESULT 2

ABL34433/c
ID ABL34433 standard; DNA: 10151 BP.XX ABL34433;
XX AC ABL34433;
XX DT 26-MAR-2002 (first entry)
XX DE Human immune system associated gene SEQ ID NO: 2406.

KW Human: immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosolic; neutropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritis; antidiabetic; antiporiatic;
KW antineoplastic; cancer; eye disease; arteriosclerotic; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
XX
XX PD
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIDEMIOLOGY AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 2406; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory ulcerative bowel
XX diseases. The present sequence is a gene of the invention
SQ
SQ Sequence 10151 BP; 3017 A; 188 C; 2298 G; 4646 T; 0 U; 2 Other;
Query Match 87.2%; Score 21.8; DB 6; Length 10151;
Best Local Similarity 92.0%; Pred. No. 50;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTGAGTA 25
Db 2331 AAAAAAAAAATCCCAATTGAGTA 2307
RESULT 3
AAS26800
ID AAS26800 standard; DNA; 12542 BP.
XX
AC AAS26800;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human genomic DNA encoding partial novel secreted protein, Seq ID 1774.
DE
XX Human; immunosuppressive; antiarthritis; ds; antirheumatic; cytosolic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
XX preservative; antiproliferative.
XX
XX Homo sapiens.
OS

XX
XX WO200155322-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001341.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 07-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 14-JUL-2000; 2000US-0218290P.
XX
XX 26-JUL-2000; 2000US-0220963P.
XX
XX 26-JUL-2000; 2000US-0220964P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
XX
XX 14-AUG-2000; 2000US-0225270P.
XX
XX 14-AUG-2000; 2000US-0225447P.
XX
XX 14-AUG-2000; 2000US-0225757P.
XX
XX 14-AUG-2000; 2000US-0225758P.
XX
XX 14-AUG-2000; 2000US-0225759P.
XX
XX 18-AUG-2000; 2000US-0226279P.
XX
XX 22-AUG-2000; 2000US-0226681P.
XX
XX 22-AUG-2000; 2000US-0226686P.
XX
XX 22-AUG-2000; 2000US-0227182P.
XX
XX 23-AUG-2000; 2000US-0227009P.
XX
XX 30-AUG-2000; 2000US-0228924P.
XX
XX 01-SEP-2000; 2000US-0229287P.
XX
XX 01-SEP-2000; 2000US-0229343P.
XX
XX 01-SEP-2000; 2000US-0229344P.
XX
XX 01-SEP-2000; 2000US-0229345P.
XX
XX 05-SEP-2000; 2000US-0229509P.
XX
XX 05-SEP-2000; 2000US-0229513P.
XX
XX 06-SEP-2000; 2000US-0230437P.
XX
XX 06-SEP-2000; 2000US-0230438P.
XX
XX 06-SEP-2000; 2000US-0231242P.
XX
XX 08-SEP-2000; 2000US-0231243P.
XX
XX 08-SEP-2000; 2000US-0231244P.
XX
XX 08-SEP-2000; 2000US-0231413P.
XX
XX 08-SEP-2000; 2000US-0231414P.
XX
XX 08-SEP-2000; 2000US-0232080P.
XX
XX 08-SEP-2000; 2000US-0232081P.
XX
XX 12-SEP-2000; 2000US-0231968P.
XX
XX 14-SEP-2000; 2000US-0232377P.
XX
XX 14-SEP-2000; 2000US-0232398P.
XX
XX 14-SEP-2000; 2000US-0232399P.
XX
XX 14-SEP-2000; 2000US-0232400P.
XX
XX 14-SEP-2000; 2000US-0232401P.
XX
XX 14-SEP-2000; 2000US-0233063P.
XX
XX 14-SEP-2000; 2000US-0233064P.
XX
XX 14-SEP-2000; 2000US-0233065P.
XX
XX 21-SEP-2000; 2000US-0234223P.
XX
XX 21-SEP-2000; 2000US-0234274P.
XX
XX 25-SEP-2000; 2000US-0234997P.
XX
XX 25-SEP-2000; 2000US-0234998P.
XX
XX 26-SEP-2000; 2000US-0235484P.
XX
XX 27-SEP-2000; 2000US-0235634P.

[illegible]

PI	Rosen CA,	Barash SC,	Ruben SM;
XX	DR	WPI; 2001-468783/53.	
XX	PT	New nucleic acid molecules encoding 461 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives.	
XX	PS	Disclosure; SEQ ID NO 1774; 980bp; English.	
CC	XX	The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISAs). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angioneurosis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence is a genomic DNA encoding a partial novel secreted protein of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format	
CC	Query Match	80.8%; Score 20.2; DB 4; Length 12542;	
CC	Best Local Similarity	88.0%; Pred. No. 2.2e+02;	
CC	Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
Oy	1	AAAAAAAAAAGTCCCAATTCGATA 25	
Dd	4239	AAAAAAAAAAAGTGCGATTTCAGATA 4263	
RESULT 4			
ID	ABX74149		
AC	ABX74149;		
XX	DT	19-MAR-2003 (first entry)	
XX	DE	Human novel polynucleotide #977.	
XX			
KW	Human; gene; ds; neural disorder; immune system disorder; renal disorder; muscular disorder; respiratory disease; reproductive disorder; gastrointestinal disorder; pulmonary disorder; cardiovascular disorder; hyperproliferative disorder; inflammatory disease; allergic reaction; blood related disorder; cancer; immunosuppressive; antiinflammatory; cardiovascular; nephrotropic; cyostatic; antiallergic; thrombolytic; haemostatic; antiarteriosclerotic.		
OS	Homo sapiens.		
XX	US2002132753-A1.		
PN	19-SEP-2002.		
PD	17-JAN-2001; 2001US-00764864.		
EF	31-JAN-2000; 2000US-0179065P.		
XX	04-FEB-2000; 2000US-0180628P.		
RR			

28-JUN-2000; 2000US-0214866P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216886P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0228686P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236357P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
PI
XX WPI; 2003-147444/14.
XX
XX New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
XX renal disorders.
XX
XX Disclosure; SEQ ID NO 1774; 402BP; English.
XX
XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage

CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
XX human novel polynucleotides of the invention
XX
SQ Sequence 12542 BP; 3802 A; 2834 C; 2642 G; 3264 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 8; Length 12542;
Best Local Similarity 88.0%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 4239 AAAAAAAAAAGTGCATTTCAGATA 4263
RESULT 5
AA166274/c
ID AA166274 standard; DNA; 13558 BP.
XX
AC AA166274;
XX
DT 22-JAN-2002 (first entry)
XX
DE Murine apoptase L100 genomic sequence.
XX
KW Apoptase; human; mouse; rat; L100; apoptosis; anticonvulsant; vasotropic;
KW neurotropic; neuroprotective; antiparkinsonian; ischemia;
KW cerebroprotective; epilepsy; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; amyotrophic lateral sclerosis; multiple sclerosis;
KW nervous system injury; stroke; ds.
XX
OS Mus musculus.
XX
PN DE10019901-A1.
XX
PD 25-OCT-2001.
XX
PF 20-APR-2000; 2000DE-01019901.
XX
PR 20-APR-2000; 2000DE-01019901.
XX
PA (BADI) BASF-LYNX BIOSCIENCE AG.
XX
XX Goetz B, Kammandel B, Kuner R, Scheek S, Hiemisch H;
PI
XX WPI; 2001-649351/75.
XX
XX New isolated nucleic acid encoding an apoptase, useful e.g. for diagnosis
PT and treatment of apoptosis-related, especially neurodegenerative,
PT diseases.
XX
XX Claim 1; Page 29-38; 78BP; German.
XX
XX The present invention provides the protein and coding sequences of
CC apoptase enzymes, designated L100, from the human, mouse and rat. The
CC protein can be used in the treatment of apoptosis related diseases,
CC including epilepsy, ischemia, Alzheimer's, Parkinson's and Huntington's
CC diseases, amyotrophic lateral sclerosis, multiple sclerosis, central
CC nervous system injury and stroke. The present sequence is the genomic
CC sequence surrounding the murine apoptase L100 coding sequence of the
XX invention
XX
SQ Sequence 13558 BP; 2936 A; 3419 C; 3606 G; 3544 T; 0 U; 53 Other;
Query Match 80.8%; Score 20.2; DB 4; Length 13558;
Best Local Similarity 88.0%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||

Db 817 AAAAAAAAAAGTCCCAATCAAGA 793

RESULT 6
AAK78660
ID AAK78660 standard; DNA; 15320 BP.
XX
AC AAK78660;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33472.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytotoxic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN W0200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205151P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0231415P.

PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234597P.
PR 25-SEP-2000; 2000US-0234598P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.

PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2000US-0259678P.
PR XX
PR (HUMA-) HUMAN GENOME SCI INC.
PR XX
PR PI Rosen CA, Baraesh SC, Ruben SM;
PR XX WPI; 2001-483426/52.
DR XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 33472; 3071bp + Sequence Listing; English.
XX
CC AAK54951 to AAK4702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 15320 BP; 4488 A; 2687 C; 3177 G; 4968 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 4; Length 15320;
Best Local Similarity 88.0%; Pred. No. 2,2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTGAGATA 25
DB 3177 AAAAAAAAAATAACAATTCAGATA 3201
RESULT 7
AAK78662
ID AAK78662 standard; DNA; 15332 BP.
XX
AC AAK78662;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33474.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; dr.
XX
OS Homo sapiens.
XX
XX WO200157182-A2.
XX
XX
PD 09-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0224513P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 16-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231422P.
PR 08-SEP-2000; 2000US-0231423P.
PR 08-SEP-2000; 2000US-0231443P.
PR 08-SEP-2000; 2000US-0231444P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235634P.
PR 27-SEP-2000; 2000US-0235636P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249267P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 XX Disclosure; SEQ ID NO 33474; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SO Sequence 15332 BP; 4490 A; 2686 C; 3180 G; 4976 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 4; Length 15332;
 Best Local Similarity 88.0%; Pred. No. 2.2e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCATTTCAGATA 25
 Db 3189 AAAAAAAAAAATACATTTCAGATA 3213
 ABL33364/C
 ID ABL33364 standard; DNA; 17893 BP.
 AC ABL33364;
 XX 26-MAR-2002 (first entry)
 DT XX
 DE Human immune system associated gene SEQ ID NO: 1337.
 XX
 XX Human; immune system disease; cytosine methylation; antiaesthetic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antihemetic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW de.
 OS XX
 OS Homo sapiens.
 PN W0200200928-A2.
 PN 03-JAN-2002.
 PD XX
 XX 02-JUL-2001; 2001MO-EP007537.
 PF 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX Claim 1; SEQ ID NO 1337; 32pp + Sequence Listing; German.
 PS The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 XX Sequence 17893 BP; 4463 A; 347 C; 4331 G; 8752 T; 0 U; 0 Other;
 SQ
 Query Match 80.8%; Score 20.2; DB 6; Length 17893;
 Best Local Similarity 88.0%; Pred. No. 2.2e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
 Db 8348 AAAAAAAAAATCCCAATTCAGATA 8324
 RESULT 9
 ACNA3994/C
 ID ACNA3994 standard; DNA; 109586 BP.
 AC ACNA3994;
 XX 18-NOV-2004 (first entry)
 DT
 XX Human genomic sequence hCG23847.
 DE
 XX
 KW Cystostatic; carcinoma; lymphoma; cancer; human; gene; ss.
 RM
 OS Homo sapiens.
 OS
 XX WO2003073826-A2.
 PN
 XX 12-SEP-2003.
 PD
 XX 28-FEB-2003; 2003WO-US06235.
 PF
 XX 01-MAR-2002; 2002US-00087192.
 PR
 XX (SAGR-) SAGRES DISCOVERY.
 PA
 PI Morris DW;
 PI
 DR WPI; 2003-328604/31.
 XX
 PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.
 PS
 XX Claim 1; SEQ ID NO 220; Opp; English.
 CC The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published
 CC
 XX Sequence 109586 BP; 31351 A; 20631 C; 21224 G; 35318 T; 0 U; 1062 Other;
 SQ

Query Match 80.8%; Score 20.2; DB 11; Length 109586;
 Best Local Similarity 88.0%; Pred. No. 2.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
 Db 97181 AAAAAAAAAAGTCCCAATTCAGATA 97157
 RESULT 10
 ABL66419/C
 ID ABL66419 standard; DNA; 348 BP.
 AC ABL66419;
 XX 15-MAY-2002 (first entry)
 DT
 XX Lung cancer related gene sequence SEQ ID NO:4756.
 DE
 XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW Cystostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 OS
 XX WO200194629-A2.
 PN
 XX 13-DEC-2001.
 PD
 XX 30-MAY-2001; 2001WO-US010838.
 PF
 XX 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233133P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 28-SEP-2000; 2000US-0236412P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.

PR 01-NOV-2000; 2000US-0245084P.
 XX (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 DR WPI; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 PS Claim 1; SEQ ID NO 4756; 44bp; English.
 XX
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms's
 CC tumour
 XX
 SQ Sequence 348 BP; 103 A; 66 C; 67 G; 110 T; 0 U; 2 Other;
 XX
 Query Match 79.2%; Score 19.8; DB 6; Length 348;
 Best Local Similarity 91.3%; Pred. No. 2.6e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
 Db 103 AAAAAATTAAGCCCAATTCAGA 81
 XX
 RESULT 11
 ID ADL13116 standard; cDNA; 588 BP.
 XX
 AC ADL13116;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Human steroid-induced C3A liver cell cDNA #845.
 XX
 KW ss; Gene; Hepatotropic; Gene therapy; Wilson disease; liver disorder;
 KW steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.
 XX
 OS Homo sapiens.
 XX
 PN US6673549-B1.
 XX
 PD 06-JAN-2004.
 XX
 PF 12-OCT-2001; 2001US-00976594.
 XX
 PR 12-OCT-2000; 2000US-0240409P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 PI Furness LM, Buchbinder JL;
 XX
 DR WPI; 2004-068610/07.
 XX

PT Combination useful for preparing a composition for treating liver
 PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,
 PT comprising cDNAs that are differentially expressed in response to steroid
 PT treatment.
 XX
 PS Claim 1; SEQ ID NO 845; 141bp; English.
 XX
 CC The invention relates to a combination comprising cDNAs that are
 CC differentially expressed in response to steroid treatment. Also included
 CC are the following: a high throughput method for using a cDNA to detect
 CC differential expression of nucleic acids in a sample; and a high
 CC throughput method of screening molecules or compounds to identify a
 CC ligand that specifically binds a cDNA. The sample is from a subject with
 CC Wilson disease and comparison of a standard defines a stage of that
 CC disease. The high throughput method of screening molecules or compounds
 CC to identify a ligand that specifically binds a cDNA comprises: combining
 CC the combination with molecules or compounds under conditions to allow
 CC specific binding; and detecting specific binding between each cDNA and at
 CC least one molecule or compound. The molecules or compounds are regulatory
 CC proteins. The combination is useful for preparing a composition for
 CC treating liver disorders associated with steroid therapy, e.g., cirrhosis
 CC or hepatitis. The present sequence represents a human cDNA which is
 CC differentially expressed in steroid-induced C3A liver cells. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 588 BP; 151 A; 152 C; 150 G; 135 T; 0 U; 0 Other;
 XX
 Query Match 79.2%; Score 19.8; DB 12; Length 588;
 Best Local Similarity 91.3%; Pred. No. 2.6e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
 Db 486 AAAAAATTAAGCCCAATTCAGA 508
 XX
 RESULT 12
 ID ADN95749 standard; DNA; 6974 BP.
 XX
 AC ADN95749;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Human BRC/LEC-related gene sequence SeqID673.
 XX
 KW growth; differentiation; blood endothelial cell; BRC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGF-R3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; angiogenesis; cytoskeletal;
 KW vasotropic; antiinflammatory; Gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; gene; ds;
 KW human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003080640-A1.
 XX
 PD 02-OCT-2003.
 XX
 PF 07-MAR-2003; 2003WO-US006900.
 XX
 PR 07-MAR-2002; 2002US-0363019P.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 XX
 PI (LICN) LICENTIA LTD.
 XX
 PI Alltalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
 XX
 DR WPI; 2003-876899/81.
 DR P-PsDB; ADN95748.
 XX

PS Example 1; SEQ ID NO 673; 176pp; English.

XX This invention relates to a method of differentially modulating the
CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
CC endothelial cells (LEC) comprising contacting endothelial cells with a
CC composition comprising an agent that differentially modulates blood or
CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
CC identifying a human subject with lymphoedema and with a mutation in at
CC least one allele of a gene encoding a LEC protein, where the mutation
CC correlates with lymphoedema in human subjects, and with the proviso that
CC the LEC protein is not VEGFR-3; and administering to the subject a
CC composition comprising a lymphatic growth agent selected from VEGF-C or
CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
CC the development of compounds with an antiangiogenic, cytostatic,
CC vasotropic or antiinflammatory activity or for gene therapy. The method
CC is useful in modulating the growth or differentiation of blood
CC endothelial cells or lymphatic endothelial cells, in treating hereditary
CC lymphoedema, in screening for an endothelial cell disorder or
CC predisposition to the disorder or in monitoring the efficacy or toxicity
CC of a drug on endothelial cells. The agent is useful in manufacturing a
CC medicament for the differential modulation of blood vessel endothelial
CC cell or lymphatic vessel endothelial cell growth or differentiation. The
CC lymphatic growth agent may also be used in manufacturing a medicament for
CC the treatment of hereditary lymphoedema resulting from a mutation in a
CC LEC gene or of other diseases involving the lymphatic vessels, such as
CC various inflammatory diseases and cancer metastasis via the lymphatic
CC system. The present sequence is that of a human LEC/BEC differentially
CC expressed gene which is related to the method of the invention. Note: This
CC sequence does not appear in the specification but was obtained by the
CC indexer using the source data given in table 14 of the specification.

SO Sequence 6974 BP; 1590 A; 1800 C; 1671 G; 1913 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 11; Length 6974;
Best Local Similarity 91.3%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGA 23
DB 2822 AAAAAACAAAGTCCCAATTAGA 2844

RESULT 13
AD019267
ID ADO19267 standard; cDNA; 6974 BP.
AC ADO19267;
XX
DT 12-AUG-2004 (first entry)
XX
DE Human PRO polynucleotide #100.

XX Human; PRO; gene; ss; immune related disorder;
KM systemic lupus erythematosus; rheumatoid arthritis; osteoarthritis;
KM juvenile chronic arthritis; systemic sclerosis; Sjogren's syndrome;
KM vasculitis; sarcoidosis; autoimmune haemolytic anaemia;
KM autoimmune thrombocytopenia; thyroiditis; diabetes mellitus;
KM renal disease; demyelinating disease; central nervous system;
KM peripheral nervous system; demyelinating polyneuropathy;
KM Guillain-Barre syndrome;
KM chronic inflammatory demyelinating polyneuropathy.

XX Homo sapiens.
OS
XX
XX WO2004043361-A2.
PN
XX
XX 27-MAY-2004.
PD
XX
XX 06-NOV-2003; 2003WO-US035268.
PF
XX 08-NOV-2002; 2002US-0425235P.
PR
XX
XX (GETH) GENENTECH INC.

XX Fong S, Dennis K, Clark H, Chiu H, Schoenfeld J, Williams PM;
PI Wood WJ, Wu TD;
PI
XX WPI; 2004-420067/39.
DR
XX P-FSDB; ADO19268.

XX Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
PT treating an immune related disorder such as systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
PT spondyloarthritis.
PS
XX Claim 1; SEQ ID NO 208; 1731pp; English.

XX The invention relates to human PRO polypeptides and the polynucleotides
CC encoding them. The polypeptides and polynucleotides are useful for
CC treating and diagnosing immune related disorders in mammals. The immune
CC related disorders include systemic lupus erythematosus, rheumatoid
CC arthritis, osteoarthritis, juvenile chronic arthritis, systemic
CC sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
CC haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
CC mellitus, immune-mediated renal disease, demyelinating diseases of the
CC central or peripheral nervous system, demyelinating polyneuropathy,
CC Guillain-Barre syndrome and chronic inflammatory demyelinating
CC polyneuropathy. This sequence represents a human PRO polynucleotide of
CC the invention.

SO Sequence 6974 BP; 1590 A; 1800 C; 1671 G; 1913 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 12; Length 6974;
Best Local Similarity 91.3%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGA 23
DB 2822 AAAAAACAAAGTCCCAATTAGA 2844

RESULT 14
AD019214
ID ADO19214 standard; DNA; 6974 BP.
XX
XX ADO19214;
AC
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 2033.
XX
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM ds.
XX
XX Homo sapiens.
OS
XX
XX WO2004048938-A2.
PN
XX
XX 10-JUN-2004.
PD
XX
XX 26-NOV-2003; 2003WO-US036193.
PF
XX 26-NOV-2002; 2002US-0429739P.
PR
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
PA
XX
XX Aziz N, Ginsburg WM, Zlotnick A;
PI
XX WPI; 2004-441208/41.
DR
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

PS Example 2; SEQ ID NO 2033; 210pp; English.

The invention relates to human tumour-associated antigenic target (TAT) polypeptides, and their related nucleic acids. The TAT polypeptides are overexpressed in cancer tissues compared to normal tissues, and may thus serve as effective targets for the diagnosis and treatment of cancer in mammals. The invention also relates to nucleic acid and polypeptide sequences at least 80% identical to the TAT nucleic acids and

compounds to identify at least one ligand that specifically binds the protein, producing or preparing polyclonal or monoclonal antibodies, or purifying antibodies from a sample. The antibodies, which specifically bind the protein differentially expressed in breast cancer is useful for detecting the expression of a protein in a sample. The BC-cDNAs are also useful for diagnosing, monitoring the treatment of, or staging, breast cancer. This sequence represents a differentially expressed breast cancer

CC associated cDNA. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from the US patent office at
CC seqdata.nefto.gov/sequence.html?docID=20020156263
XX
SQ Sequence 7032 BP; 1569 A; 1816 C; 1703 G; 1906 T; 0 U; 38 Other;
Query Match 79.2%; Score 19.8; DB 8; Length 7032;
Best Local Similarity 91.3%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAGTCCCAATTGAGA 23
Db 2875 AAAAAACAAGTCCCAATTGAGA 2897
RESULT 17
ADQ23503
ID ADQ23503 standard; DNA; 7835 BP.
XX
AC ADQ23503;
XX
XX 26-AUG-2004 (first entry)
XX
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 6323.
XX
XX soft tissue sarcoma; cytoskeletal; gene therapy; vaccine; screening; human;
XX ds.
XX Homo sapiens.
XX
XX WO2004048938-A2.
XX
XX 10-JUN-2004.
XX
XX 26-NOV-2003; 2003WO-US038193.
XX
XX 26-NOV-2002; 2002US-0429739P.
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
XX
XX Aziz N, Ginsburg WM, Zlotnick A;
XX WPI; 2004-441208/41.
XX
XX Bearly detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.
XX
XX Example 2; SEQ ID NO 6323; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytoskeletal applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 7835 BP; 1789 A; 1990 C; 1979 G; 1993 T; 0 U; 84 Other;
Query Match 79.2%; Score 19.8; DB 12; Length 7835;
Best Local Similarity 91.3%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAGTCCCAATTGAGA 23

Db 2896 AAAAAACAAGTCCCAATTGAGA 2918
RESULT 18
AAL37105
ID AAL37105 standard; DNA; 10514 BP.
XX
XX AAL37105;
XX
XX 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3470.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antifungal;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.
XX
XX Homo sapiens.
XX
XX WO200155367-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001338.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205151P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 07-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 11-JUL-2000; 2000US-0217496P.
XX
XX 14-JUL-2000; 2000US-0218290P.
XX
XX 26-JUL-2000; 2000US-0220963P.
XX
XX 26-JUL-2000; 2000US-0220964P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
XX
XX 14-AUG-2000; 2000US-0225270P.
XX
XX 14-AUG-2000; 2000US-0225447P.
XX
XX 14-AUG-2000; 2000US-0225757P.
XX
XX 14-AUG-2000; 2000US-0225758P.
XX
XX 14-AUG-2000; 2000US-0225759P.
XX
XX 18-AUG-2000; 2000US-0226279P.
XX
XX 22-AUG-2000; 2000US-0226681P.
XX
XX 22-AUG-2000; 2000US-0226688P.
XX
XX 23-AUG-2000; 2000US-0227182P.
XX
XX 23-AUG-2000; 2000US-0227009P.
XX
XX 30-SEP-2000; 2000US-0229282P.
XX
XX 01-SEP-2000; 2000US-0229287P.
XX
XX 01-SEP-2000; 2000US-0229343P.
XX
XX 01-SEP-2000; 2000US-0229344P.
XX
XX 01-SEP-2000; 2000US-0229345P.
XX
XX 05-SEP-2000; 2000US-0229509P.
XX
XX 05-SEP-2000; 2000US-0229513P.
XX
XX 06-SEP-2000; 2000US-0230437P.
XX
XX 06-SEP-2000; 2000US-0230438P.

```

PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232357P.
PR 14-SEP-2000; 2000US-0232358P.
PR 14-SEP-2000; 2000US-0232359P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236378P.
PR 29-SEP-2000; 2000US-0236388P.
PR 29-SEP-2000; 2000US-0236389P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

```

```

PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251888P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-451937/48.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX disorders related to the musculoskeletal system including musculoskeletal
XX cancers and also for testing and detection e.g. diagnosis.
XX
XX Example 2; SEQ ID NO 3470; 781pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
XX (AAB03087-AB04109) associated with the musculoskeletal system useful for
XX preventing, treating or ameliorating medical conditions e.g. by protein
XX or gene therapy. The genes are isolated from a range of human tissues
XX disclosed in the specification. The nucleic acids, proteins, antibodies
XX and (ant)agonists are useful in the diagnosis, treatment and prevention
XX of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
XX adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
XX lung, or urogenital; (b) immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; (c) cardiovascular disorders such as
XX myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
XX cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
XX bacterial, fungal and parasitic infections. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 10514 BP; 2323 A; 2744 C; 2655 G; 2792 T; 0 U; 0 Other;
XX
XX Query Match 79.2%; Score 19.8; DB 4; Length 10514;
XX Best Local Similarity 91.3%; Pred. No. 3; 1e+02;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCATTGAGA 23
Db 6354 AAAAAACAAAAGTCCATTGAGA 6376
RESULT 19
ABX60093
ID ABX60093 standard; cDNA; 10514 BP.
XX
XX ABX60093;
AC ABX60093;
XX
XX 26-FEB-2003 (first entry)
XX
XX cDNA encoding novel human musculoskeletal system antigen #2437.
DE
XX
XX Gene; ss; musculoskeletal system antigen; cancer; metastasis;
KW

```


XX Homo sapiens.
OS
XX
XX US2004009488-A1.
XX
PD 15-JAN-2004.
XX
PF 13-SEP-2002; 2002US-00242515.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231966P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246539P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251888P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 11-DEC-2000; 2000US-0254097P.
PR 17-JAN-2001; 2001US-00764877.

XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Barash SC;
XX WPI; 2004-090458/09.
XX
XX New nucleic acid molecule, useful for preparing a medicament for
XX preventing, treating or ameliorating a medical condition e.g., cancer of
XX musculoskeletal tissues or osteoporosis.
XX
XX Disclosure; SEQ ID NO 3470; 289pp; English.
XX
XX The invention relates to a novel isolated musculoskeletal system-
XX associated nucleic acid molecule. The nucleic acid of the invention
XX demonstrates cytostatic and osteopathic activities and may be useful for
XX preparing a medicament for preventing, treating or ameliorating a medical
XX condition such as cancer of the musculoskeletal tissues or osteoporosis,
XX possibly via gene therapy or vaccine production. The current sequence is
XX that of the human musculoskeletal system-associated genomic DNA of the
XX invention. The current sequence is not shown within the specification per
XX se but is available on the USPTO web-site
XX http://seqdata.uspto.gov/sequence.html?DocID=20040009488.
XX
XX Sequence 10514 BP; 2323 A; 2744 C; 2655 G; 2792 T; 0 U; 0 Other;
XX
XX Query Match 79.2%; Score 19.8; DB 12; Length 10514;
XX Best Local Similarity 91.3%; Pred. No. 3.1e+02;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAGTCCCAATTGAGA 23
XX Db 6354 AAAAAACAAAGTCCCAATTGAGA 6376
XX
XX RESULT 21
XX ADT18127
XX ID ADT18127 standard; cDNA; 991 BP.
XX AC ADT18127;
XX DT 13-JAN-2005 (first entry)
XX DE Plant cDNA, Seq ID 3453.
XX
XX Plant; ss; gene; transgenic; cold tolerance; growth rate;
XX drought tolerance; disease resistance; galactomannan production;
XX plant growth regulator; heat tolerance; herbicide tolerance;
XX lignin production; extreme osmotic condition tolerance;
XX pathogens resistance; pest resistance; yield improvement; seed oil yield;
XX seed protein yield.
XX
XX Vairidiplantae.
XX OS
XX US2004216190-A1.
XX
XX 28-OCT-2004.
XX
XX 18-DEC-2003; 2003US-00739930.
XX
XX 28-APR-2003; 2003US-00424599.
XX
XX 28-APR-2003; 2003US-00425115.
XX
XX (KOVA/) KOVALIC D K.
XX
XX KOVALIC DK;
XX
XX WPI; 2004-757369/74.
XX
XX New recombinant DNA constructs useful in the field of biochemistry and
XX genetics, and in particular for producing transgenic plants with improved
XX biological characteristics.

PS Claim 1; SEQ ID NO 3453; 14pp; English.
XX
XX The invention relates a recombinant DNA construct comprising a
XX polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:
XX 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
XX (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,
XX Arabidopsis, wheat and rape but the specification does not indicate which
XX sequences is derived from which organism. Also included is a method of
XX producing a plant having an improved property, comprising transforming a
XX plant with a recombinant DNA construct comprising a promoter region
XX functional in a plant cell operably joined to a polynucleotide encoding a
XX polypeptide associated with the property, and growing the transformed
XX plant. The property is selected from improving plant cold tolerance, for
XX manipulating growth rate in plant cells by modification of the cell cycle
XX pathway, for improving plant drought tolerance, for providing increased
XX resistance to plant disease, for galactomannan production, for production
XX of plant growth regulators, for improving plant heat tolerance, for
XX improving plant tolerance to herbicides, for increasing the rate of
XX homologous recombination in plants, for lignin production, for improving
XX plant tolerance to extreme osmotic conditions, for improving plant
XX tolerance to pathogens or pests, for yield improvement by modification of
XX photosynthesis, for modifying seed oil yield and/or content, for
XX modifying seed protein yield and/or content, for yield improvement by
XX modification of carbohydrate, nitrogen or phosphorus use and/or uptake
XX and for yield improvement by providing improved plant growth and
XX development under at least one stress condition. The polynucleotide may
XX also encode a plant transcription factor. The methods and compositions of
XX the present invention are useful in the field of biochemistry and
XX genetics, in particular for producing transgenic plants with improved
XX biological characteristics such as increased yield, improved nitrogen
XX flow, increasing plant tolerance to cold or heat, improving plant
XX tolerance to extreme osmotic and drought conditions, and improving plant
XX tolerance to plant pests or pathogens. They can also be used in physical
XX arrays of molecules, plant breeding markers, computer-based storage and
XX analysis systems. The present sequence is one of the 5544 plant cDNA
XX sequences of the invention. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20040216190.
XX
XX Sequence 991 BP; 303 A; 215 C; 194 G; 279 T; 0 U; 0 Other;
XX
XX Query Match 77.6%; Score 19.4; DB 13; Length 991;
XX Best Local Similarity 95.2%; Pred. No. 3.9e+02;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAGTCCCAATTGCA 21
XX Db 24 AAAAAAAAAAATCCCAATTGCA 44
XX
XX RESULT 22
XX AAZ60888/C
XX ID AAZ60888 standard; DNA; 17131 BP.
XX AC AAZ60888;
XX
XX 16-MAY-2000 (first entry)
XX
XX DNA encoding a human geranylgeranyl pyrophosphate synthetase (hGGPPS).
XX
XX Human; geranylgeranyl pyrophosphate synthetase; hGGPPS; chromosome 1;
XX 1q42-1q43 locus; prostate cancer; hGGPPS; diallelic marker;
XX mevalonic biosynthetic pathway; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 486..546
XX exon /*tag= a
XX FT /number= 1
XX FT 547..7291
XX FT intron /*tag= b


```

XX DE Hereditary haemochromatosis subregion from an unaffected individual.
XX KM
XX KM Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
XX KM diagnosis; iron metabolism; NPT3; NPT4; Roret; BTFL; BTF2; BTF3; BTF4;
XX KM BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
XX KM type 1 sodium transport gene; ss.
XX OS
XX OS Homo sapiens.
XX PN
XX PN MO9814466-A1.
XX PD
XX PD 09-APR-1998.
XX PF
XX PF 30-SEP-1997; 97WO-US017658.
XX PR
XX PR 01-OCT-1996; 96US-00724394.
XX PR 07-MAY-1997; 97US-00852495.
XX XX
XX XX (PROG-) PROGENITOR INC.
XX PA
XX PA Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ;
XX PI Teuchshaehi Z, Wolff RK;
XX DR WPI; 1998-240014/21.
XX XX
XX PT Hereditary haemochromatosis gene products - used to develop products for
XX PT the diagnosis and treatment of hereditary disorders in iron metabolism.
XX PS
XX PS Example 2; Fig 8; 209pp; English.
XX XX
XX CC The present invention describes hereditary haemochromatosis gene products
XX CC from the human haemochromatosis gene. The present sequence represents a
XX CC hereditary haemochromatosis subregion from an individual unaffected by
XX CC hereditary haemochromatosis (HH). Also described is a method to determine
XX CC the presence or absence of the common hereditary haemochromatosis (HFE)
XX CC gene mutation in an individual comprising: (a) providing DNA or RNA from
XX CC the individual; and (b) assessing the DNA or RNA for the presence or
XX CC absence of a haplotype or genotype where the presence or absence of the
XX CC haplotype genotype indicates the likely presence of the HFE gene mutation
XX CC in the genome of the individual. The HFE gene sequences from the present
XX CC invention can be used to develop products for use in the diagnosis and
XX CC treatment of HFE. The present invention also describes BTF genes, which
XX CC are homologues of the milk protein butyrophilin (BT), and can be used in
XX CC the production of agonists and antagonists of BT function. Also described
XX CC are: (1) a Roret gene which can be used to develop products for the
XX CC study, diagnosis and treatment of lupus and Sjogren's syndrome; and (2)
XX CC NPT3 and NPT4 genes which are homologues of a type 1 sodium transport
XX CC gene, and can similarly be used for hypophosphatemia
XX SO
XX SO Sequence 235033 BP; 68786 A; 48466 C; 49441 G; 68340 T; 0 U; 0 Other;
XX
XX Query Match 77.6%; Score 19.4; DB 2; Length 235033;
XX Best Local Similarity 95.2%; Pred. No. 5.4e+02;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAAGTCCCAATTCA 21
XX Db 187050 AAAAAAAAAAGTCCCAATTCA 187030
XX
XX RESULT 25
XX ID AAV57903 standard; DNA; 237326 BP.
XX AC AAV57903;
XX XX
XX XX 21-DEC-1998 (first entry)
XX DE Hereditary haemochromatosis subregion from an HH affected individual.
XX KM Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
XX KM diagnosis; iron metabolism; NPT3; NPT4; Roret; BTFL; BTF2; BTF3; BTF4;
XX KM

```

```

XX KM BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
XX KM type 1 sodium transport gene; ss.
XX OS
XX OS Homo sapiens.
XX PN
XX PN MO9814466-A1.
XX PD
XX PD 09-APR-1998.
XX PF
XX PF 30-SEP-1997; 97WO-US017658.
XX PR
XX PR 01-OCT-1996; 96US-00724394.
XX PR 07-MAY-1997; 97US-00852495.
XX XX
XX XX (PROG-) PROGENITOR INC.
XX PA
XX PA Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ;
XX PI Teuchshaehi Z, Wolff RK;
XX DR WPI; 1998-240014/21.
XX XX
XX PT Hereditary haemochromatosis gene products - used to develop products for
XX PT the diagnosis and treatment of hereditary disorders in iron metabolism.
XX PS
XX PS Claim 1; Fig 9; 209pp; English.
XX XX
XX CC The present invention describes hereditary haemochromatosis gene products
XX CC from the human haemochromatosis gene. The present sequence represents a
XX CC hereditary haemochromatosis subregion from an hereditary haemochromatosis
XX CC (HH) affected individual. Also described is a method to determine the
XX CC presence or absence of the common hereditary haemochromatosis (HFE) gene
XX CC mutation in an individual comprising: (a) providing DNA or RNA from the
XX CC individual; and (b) assessing the DNA or RNA for the presence or absence
XX CC of a haplotype or genotype where the presence or absence of the haplotype
XX CC genotype indicates the likely presence of the HFE gene mutation in the
XX CC genome of the individual. The HFE gene sequences from the present
XX CC invention can be used to develop products for use in the diagnosis and
XX CC treatment of HFE. The present invention also describes BTF genes, which
XX CC are homologues of the milk protein butyrophilin (BT), and can be used in
XX CC the production of agonists and antagonists of BT function. Also described
XX CC are: (1) a Roret gene which can be used to develop products for the
XX CC study, diagnosis and treatment of lupus and Sjogren's syndrome; and (2)
XX CC NPT3 and NPT4 genes which are homologues of a type 1 sodium transport
XX CC gene, and can similarly be used for hypophosphatemia
XX SO
XX SO Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T; 0 U; 0 Other;
XX
XX Query Match 77.6%; Score 19.4; DB 2; Length 237326;
XX Best Local Similarity 95.2%; Pred. No. 5.4e+02;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAAGTCCCAATTCA 21
XX Db 189274 AAAAAAAAAAGTCCCAATTCA 189254
XX
XX RESULT 26
XX ID AEB67455
XX AC AEB67455;
XX XX
XX XX 22-SEP-2005 (first entry)
XX DE Rice genome derived DNA sequence, SEQ ID 2600.
XX KM transfection; gene regulation; transgenic plant; RNA interference;
XX KM transformation; antibody; de.
XX OS
XX OS Oryza sp.
XX XX
XX XX JP2005185101-A.
XX XX

```


[illegible]

```
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0234984P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236357P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239377P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246479P.
PR 08-NOV-2000; 2000US-0246479P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249256P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251889P.
PR 08-DEC-2000; 2000US-0251900P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 41957; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytotoxic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention
XX
```


Sequence 978 BP, 316 A; 154 C; 179 G; 329 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 4; Length 978;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 247 AAAAAAAAAAGTCCCAATTCAGAT 270
RESULT 29
AAK87147
ID AAK87147 standard; DNA; 1397 BP.
XX
AC AAK87147;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41959.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytosclatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN NO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198113P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216860P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0231415P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.

XX	PR	17-NOV-2000;	2000US-0249217P.
XX	PR	17-NOV-2000;	2000US-0249218P.
XX	PR	17-NOV-2000;	2000US-0249219P.
XX	PR	17-NOV-2000;	2000US-0249244P.
XX	PR	17-NOV-2000;	2000US-0249245P.
XX	PR	17-NOV-2000;	2000US-0249246P.
XX	PR	17-NOV-2000;	2000US-0249265P.
XX	PR	17-NOV-2000;	2000US-0249297P.
XX	PR	17-NOV-2000;	2000US-0249299P.
XX	PR	17-NOV-2000;	2000US-0249300P.
XX	PR	01-DEC-2000;	2000US-0250160P.
XX	PR	01-DEC-2000;	2000US-0250391P.
XX	PR	05-DEC-2000;	2000US-0251030P.
XX	PR	05-DEC-2000;	2000US-0251988P.
XX	PR	05-DEC-2000;	2000US-0251989P.
XX	PR	06-DEC-2000;	2000US-0251479P.
XX	PR	08-DEC-2000;	2000US-0251856P.
XX	PR	08-DEC-2000;	2000US-0251868P.
XX	PR	08-DEC-2000;	2000US-0251869P.
XX	PR	08-DEC-2000;	2000US-0251989P.
XX	PR	08-DEC-2000;	2000US-0251990P.
XX	PR	11-DEC-2000;	2000US-0254097P.
XX	PR	05-JAN-2001;	2001US-0259678P.
XX	PA	(HUMA-)	HUMAN GENOME SCT INC.
XX	XX		
XX	P1	Rosen CA, Barash SC, Ruben SM;	
XX	DR	WPI; 2001-483426/52.	
XX	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
XX	XX	useful for preventing, diagnosing and/or treating cancers and metastasis.	
XX	PS	Disclosure; SEQ ID NO 41959; 3071pp + Sequence Listing; English.	
XX	CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)	
XX	CC	amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic	
XX	CC	activity, and can be used in gene therapy and vaccine production. (I)	
XX	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
XX	CC	treatment of diseases associated with inappropriate (I) expression. For	
XX	CC	example, they may be used to treat disorders associated with decreased	
XX	CC	expression by rectifying mutations or deletions in a patient's genome	
XX	CC	that affect the activity of (I) by expressing inactive proteins or to	
XX	CC	supplement the patient's own production of (I). Additionally, (I)	
XX	CC	polynucleotides may be used to produce the secreted (I), by inserting the	
XX	CC	nucleic acids into a host cell and culturing the cell to express the	
XX	CC	protein. (I) proteins and polynucleotides may be used to prevent,	
XX	CC	diagnose and treat immune/hematopoietic-related diseases, especially	
XX	CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703	
XX	CC	to AAK67694 represent human immune/hematopoietic antigen genomic	
XX	CC	sequences from the present invention. AAK54942 to AAK54950 and AAK62169	
XX	CC	represent sequences used in the exemplification of the present invention	
XX	SO	Sequence 1397 BP; 438 A; 212 C; 272 G; 475 T; 0 U; 0 Other;	
OY	Query Match	76.8%; Score 19.2; DB 4; Length 1397;	
DB	Best Local Similarity	87.5%; Pred. No. 4.8e+02;	
DB	Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
OY	1 AAAAAAAAAAGTCCCAATTCAGAT 24		
DB	666 AAAAAAAAAAGTCCCACTTACAT 689		
XX	RESULT 30		
XX	ID ABAZ70778 standard; CDNA; 1821 BP.		
XX	ABZ70778;		
XX	ABZ70778;		
XX	14-APR-2003 (first entry)		
XX	Human Parkinson's disease associated protein 20.13-encoding cDNA.		

KW	Human; Parkinson's disease associated protein 20.13;
KM	recombinant production; gene therapy; cancer; tumour; HIV infection;
KV	human immunodeficiency virus; cytosstatic; gene; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	2..553
FT	/*cag= a
FT	/product= "Human Parkinson's disease associated protein
FT	20.13"
XX	
PN	CN1359949-A.
XX	
PD	24-JUL-2002.
XX	
PF	20-DEC-2000; 2000CN-00135166.
XX	
PR	20-DEC-2000; 2000CN-00135166.
XX	
PA	(BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX	
PZ	Mao Y, Xie Y;
XX	
DR	WPI; 2002-733637/80.
DR	P-PSDB; ABP58960.
XX	
PT	Polypeptide-human Parkinson syndrome related protein 20.13 and
PT	polynucleotide for coding it.
XX	
PS	Claim 6; Page 26-27 (Disclosure); 34pp; Chinese.
XX	
CC	The invention relates to human Parkinson's disease associated protein
CC	20.13 (ABP58960) and nucleic acids encoding it (ABZ70778). The protein
CC	has a molecular weight of 20.13 kD. The invention also relates to a
CC	method for the recombinant production of the protein, an antagonist of
CC	the protein, and the use of the protein, gene and antagonist in
CC	therapeutic applications. Parkinson's disease associated protein 20.13
CC	can be used in the treatment of a variety of diseases such as cancer and
CC	HIV (human immunodeficiency virus) infection. The present sequence
CC	represents cDNA encoding human Parkinson's disease associated protein
CC	20.13
SQ	Sequence 1821 BP; 553 A; 301 C; 340 G; 627 T; 0 U; 0 Other;
XX	
Query Match	76.8%; Score 19.2; DB 6; Length 1821;
Best Local Similarity	87.5%; Pred. No. 4.9e+02;
Matches 21; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
OY	1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db	1096 AAAAAAAAAAGTCCCACTTACAT 1119
RESULT 31	
ADXS5035/c	
ID ADXS5035 standard; cDNA; 2567 BP.	
XX	
AC ADXS5035;	
XX	
DT 21-APR-2005 (first entry)	
XX	
DE Plant full length insert polynucleotide seqid 29775.	
XX	
KM plant protectant; plant growth regulant; gene therapy; plant;	
KM recombinant DNA construct; physical array; plant breeding marker;	
KM cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;	
KM extreme osmotic condition; pathogen tolerance; pest tolerance;	
KM growth rate; cell cycle pathway; disease resistance;	
KM galactomannan production; lignin production; plant growth regulator;	
KM yield; plant growth; plant development; seed oil; protein yield;	
KM protein content; gene; ss.	
XX	

```

OS Unidentified.
XX
XX US200403488-A1.
XX
XX 19-FEB-2004.
XX
XX 28-APR-2003; 2003US-00425114.
XX
XX 06-MAY-1999; 99US-00304517.
XX
XX 05-NOV-2001; 2001US-00985678.
XX
XX (LIU/) LIU J.
XX (ZHOU/) ZHOU Y.
XX (KOVA/) KOVALIC D K.
XX (SCRE/) SCREEN S E.
XX (TABA/) TABASKA J E.
XX (CAOY/) CAO Y.
XX
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
XX pests, for conferring increased resistance to plant disease, or for
XX improving yield.
XX
XX Claim 1; SEQ ID NO 29775; 15pp; English.
XX
XX The invention describes a recombinant DNA construct comprising a
XX polynucleotide consisting of a sequence encoding an amino acid sequence
XX available in electronic form from the US patent office at
XX ftp.segdata.uspto.gov/sequence.html?DocID:200403488. The polynucleotide
XX of the invention are also useful in physical arrays of molecules and as
XX plant breeding markers. The recombinant DNA construct is useful for
XX improving plant tolerance to cold, heat, drought, herbicides, extreme
XX osmotic conditions, pathogens or pests, for manipulating growth rate in
XX plant cells by modification of the cell cycle pathway, for conferring
XX increased resistance to plant disease, for producing galactomannan,
XX lignin or plant growth regulators, for increasing the rate of homologous
XX recombination in plants, for improving yield by modification of
XX photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
XX or by providing improved plant growth and development under at least one
XX stress condition or for modifying seed oil or protein yield and/or
XX content. This sequence represents a plant full length insert
XX polynucleotide that can be used in the recombinant DNA construct of the
XX invention.
XX
XX Sequence 2567 BP; 699 A; 461 C; 524 G; 883 T; 0 U; 0 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 13; Length 2567;
XX Best Local Similarity 87.5%; Pred. No. 5e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 2 AAAAAAAAAAGTCCCAATTGAGTA 25
XX Db 2388 AAAAAAAAAAGTGAATTGAGTA 2365
XX
XX RESULT 32
XX ADQ64807
XX ID ADQ64807 standard; cDNA; 2690 BP.
XX
XX AC ADQ64807;
XX
XX DT 07-OCT-2004 (first entry)
XX
XX XX Novel human cDNA sequence #1968.
XX
XX se; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
XX cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
XX neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
XX cancer.

```

```

XX
XX Homo sapiens.
OS
XX
XX EPI440981-A2.
XX
XX 28-JUL-2004.
XX
XX 21-JAN-2004; 2004EP-00001196.
XX
XX 21-JAN-2003; 2003JP-00102206.
XX
XX 09-MAY-2003; 2003JP-00131392.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX Yamamoto J, Isono Y, Nagai K, Irie R;
XX WPI; 2004-535376/52.
XX
XX P-PSDB; ADQ66995.
XX
XX Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
XX Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
XX Claim 1; SEQ ID NO 1968; 2449pp; English.
XX
XX The invention relates to 2495 novel polynucleotides (I) and their encoded
XX polypeptides, sequences hybridizing to these nucleotides, sequences
XX encoding partial polypeptides and sequences having 70% or 90% identity to
XX the nucleotide and protein sequences. The nucleotides and polypeptides
XX are useful as diagnostic markers or therapeutic target for the diseases
XX or morbid states. They are also useful for treating osteoporosis,
XX neurological diseases, Alzheimer's diseases, Parkinson's diseases,
XX dementia and various cancers. This sequence corresponds to a nucleotide
XX sequence of the invention.
XX
XX Sequence 2690 BP; 826 A; 457 C; 516 G; 891 T; 0 U; 0 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 12; Length 2690;
XX Best Local Similarity 87.5%; Pred. No. 5e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAGTCCCAATTGAGT 24
XX Db 1961 AAAAAAAAAAGTCCCAATTGAGT 1984
XX
XX RESULT 33
XX ADH69515/C
XX ID ADH69515 standard; DNA; 4411 BP.
XX
XX AC ADH69515;
XX
XX DT 25-MAR-2004 (first entry)
XX
XX DE Pig dopamine D1 receptor DNA.
XX
XX XX Skeletal muscle; dopamine receptor; therapy; muscle atrophy; hypertrophy;
XX muscular dystrophy; gene; de; pig.
XX
XX OS Sus scrofa.
XX
XX Key Location/Qualifiers
XX CDS 1708..3048
XX FT /*tag= a
XX FT /product= "Pig dopamine D1 receptor"
XX
XX PN US2003170741-A1.
XX
XX PD 11-SEP-2003.
XX
XX 18-NOV-2002; 2002US-00299642.
XX
XX 18-JAN-2002; 2002US-0349620P.

```

XX (PROC) PROCTER & GAMBLE CO.
 PA Isfort RJ, Sheldon RJ;
 PI WPI; 2004-069018/07.
 DR P-PSDB; ADH69516.
 XX
 PT Identifying a compound for regulating activity or expression of a
 PT dopamine receptor, useful in identifying compounds to regulate skeletal
 PT muscle mass or function, to treat muscle atrophy, induce hypertrophy and
 PT treat muscular dystrophies.
 XX
 PS Disclosure; SEQ ID NO 31; 74pp; English.
 CC The invention relates to a method for identifying candidate compounds for
 CC regulating skeletal muscle mass or function. The method comprising
 CC contacting a test compound with a D1 or D5 dopamine receptor and
 CC identifying those compounds that bind to or activate the receptor. The
 CC method is used to regulate skeletal muscle mass or function, to treat
 CC muscle atrophy or to induce hypertrophy or treat muscular dystrophies.
 CC The present sequence is pig dopamine D1 receptor DNA.
 XX
 SQ Sequence 4411 BP; 993 A; 1140 C; 1171 G; 1107 T; 0 U; 0 Other;
 Query Match 76.8%; Score 19.2; DB 12; Length 4411;
 Best Local Similarity 87.5%; Pred. No. 5.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
 DB 4321 AAAAAAAAAATCTCAGTTCAAT 4298
 RESULT 34
 AAL44712
 ID AAL44712 standard; DNA; 5097 BP.
 AC AAL44712;
 DT 03-MAY-2002 (first entry)
 XX
 DE Human glutamate receptor delta-1 subunit coding sequence SEQ ID NO: 12.
 XX
 KW Human; glutamate receptor delta-1 subunit; epilepsy; brain damage;
 KW neurodegenerative disorder; Huntington's disease; Parkinson's disease;
 KW Alzheimer's disease; schizophrenia; mood disorder; dementia;
 KW neuropathological pain; pain; receptor; anticonvulsant; noctropic;
 KW neuroprotective; vasotropic; analgesic; neuroleptic; cyrostatic;
 KW utropathic; antiparkinsonian; vulnerary; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200206313-A2.
 XX
 PD 24-JAN-2002.
 XX
 PF 13-JUL-2001; 2001WO-EP008102.
 XX
 PR 18-JUL-2000; 2000US-0218835P.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Kossida S;
 XX
 DR WPI; 2002-195800/25.
 XX
 PT Novel human glutamate receptor delta-1 subunit protein which can be
 PT regulated for treating epilepsy, schizophrenia, neurodegenerative
 PT diseases, ischemia, pain, benign prostate hyperplasia and urinary
 PT incontinence.
 XX
 PS Claim 19; Fig 15; 97pp; English.

XX The present invention provides a human glutamate receptor delta-1 subunit
 CC polypeptide. This can be used to screen for agents which modulate the
 CC activity of glutamate receptor delta-1 subunit polypeptide, which may
 CC then be used in the treatment of diseases such as epilepsy, brain damage,
 CC neurodegenerative disorders such as Alzheimer's disease, Huntington's
 CC disease and Parkinson's disease, schizophrenia, mood disorder, pain,
 CC neuropathologic pain and dementia. The present sequence is the human
 CC glutamate receptor delta-1 subunit coding sequence
 XX
 SQ Sequence 5097 BP; 1236 A; 1291 C; 1356 G; 1214 T; 0 U; 0 Other;
 Query Match 76.8%; Score 19.2; DB 6; Length 5097;
 Best Local Similarity 87.5%; Pred. No. 5.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
 DB 4924 AAAAAAAAACTCCCACTCAGAT 4947
 RESULT 35
 AAC74570
 ID AAC74570 standard; cDNA; 5503 BP.
 AC AAC74570;
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF125 polynucleotide sequence SEQ ID NO:249.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerary; antiparinsonian; noctropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiallergic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antihemantic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; inflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 XX
 DR P-PSDB; AAB40361.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 5; Page 560-564; 5507pp; English.
 XX
 CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipneumatic; antipneumonia; nocrotropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antichryoid; and antanaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease, to
CC enhance coagulation, to inhibit thrombosis, and as a contraceptive
CC XX

Sequence 5503 BP; 1322 A; 1420 C; 1463 G; 1298 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 3; Length 5503;
Best Local Similarity 87.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
5330 AAAAAAAAAACTCCCAACTCGAT 5353

Db

RESULT 36
AAS36240
ID AAS36240 standard; DNA; 5628 BP.
XX
AC AAS36240;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen genomic DNA SEQ ID No 1740.
XX
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nocrotropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infectivity.
XX
XX Homo sapiens.
XX
XX PN MO200155321-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001MO-US001340.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225265P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.

CC	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251899P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA,	Barash SC,
XX		Ruben SM;
XX		
DR	WPI;	2001-451930/48.
XX		
PT	New cardiovascular system related polynucleotides and polypeptides,	
PT	useful for diagnosing,	treating and/or preventing disorders of the
XX	cardiovascular system.	
XX		
PS	Claim 1,	SEQ ID NO 1740; 674dp; English.
XX		
CC	Sequences AAS3741-AAS3642 represent genomic DNA molecules, which encoded	
CC	the cardiovascular system antigen polypeptides of the invention.	
CC	Cardiovascular system antigens and their associated polynucleotides are	
CC	useful in the diagnosis, treatment and prevention of various types of	
CC	disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,	
CC	chickens or sheep. A pathological condition can be determined by	
CC	detecting the presence or absence of a mutation in a cardiovascular	
CC	system antigen polynucleotide. The treatable disorders include autoimmune	
CC	diseases such as rheumatoid arthritis, hyperproliferative disorders such	

CC	as neoplasms of the breast or liver, cardiovascular disorders such as
CC	cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC	nervous system disorders such as Alzheimer's disease, infections caused
CC	by bacteria, viruses and fungi, ocular disorders such as corneal
CC	infection, endocrine disorders such as premature labour and infertility,
CC	gastrointestinal disorders such as Crohn's disease, renal disorders such
CC	as glomerulonephritis and respiratory disorders such as asthma and
CC	pleurisy. The polypeptides can also be used to aid wound healing, to
CC	prevent skin aging due to sunburn, to maintain organs before
CC	transplantation, to regenerate tissues and in chemotaxis. Note: The
CC	sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences
XX	
Query Match	76.8%; Score 19.2; DB 4; Length 5628;
Best Local Similarity	87.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
QY	1 AAAAAAAAAAGTCCCAATTCAGAT 24
DB	4860 AAAAAAAAAATCCCTATTGTGAT 4883
RESULT 37	
AL03224	
ID	AL03224 standard; DNA; 5628 BP.
XX	
AC	AL03224;
XX	
XX	21-NOV-2001 (first entry)
DT	
XX	
DE	Human reproductive system related antigen DNA SEQ ID NO: 5912.
XX	
KW	Human: reproductive system related antigen; reproductive system disorder;
KW	cancer; gene therapy; ds.
XX	
OS	Homo sapiens.
XX	
XX	WO200155320-A2.
PD	
XX	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US001339.
XX	
PR	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-018464P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205515P.
PR	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214886P.
PR	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.
PR	07-JUL-2000; 2000US-0216880P.
PR	11-JUL-2000; 2000US-0217487P.
PR	11-JUL-2000; 2000US-0217496P.
PR	14-JUL-2000; 2000US-0218290P.
PR	26-JUL-2000; 2000US-0220963P.
PR	14-AUG-2000; 2000US-0224518P.
PR	14-AUG-2000; 2000US-0224519P.
PR	14-AUG-2000; 2000US-0225213P.
PR	14-AUG-2000; 2000US-0225214P.
PR	14-AUG-2000; 2000US-0225266P.
PR	14-AUG-2000; 2000US-0225267P.
PR	14-AUG-2000; 2000US-0225268P.
PR	14-AUG-2000; 2000US-0225270P.
PR	14-AUG-2000; 2000US-0225447P.
PR	14-AUG-2000; 2000US-0225475P.

```
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 19-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228992P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239357P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241121P.
PR 20-OCT-2000; 2000US-0241178P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX
XX DR WPI; 2001-465570/50.
XX
XX PT Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX PS Disclosure; SEQ ID NO 5912; 1297bp + Sequence listing; English.
XX
XX CC The present invention provides the protein and coding sequences of a
XX CC number of human reproductive system related antigens. These can be used
XX CC in the prevention and treatment of reproductive system disorders,
XX CC including cancer. The present sequence is a genomic sequence encoding a
XX CC protein of the invention
XX
XX SQ Sequence 5628 BP; 1449 A; 1395 C; 1432 G; 1352 T; 0 U; 0 Other;
XX

Query Match 76.8%; Score 19.2; DB 4; Length 5628;
Best Local Similarity 87.5%; Pred. No. 5 2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAGAT 24
Db 4860 AAAAAAAAAATCCCTATTAGAT 4883

RESULT 38
ADE46934
ID ADE46934 standard; DNA; 5628 BP.
XX
XX ADE46934;
AC
XX
XX 29-JAN-2004 (first entry)
DT
XX
XX Human cardiovascular system related genomic DNA #500.
DE
```

XX Human; cardiovascular system related polypeptide; cancer;
KW Proliferative disorder; foetal abnormality; developmental abnormality;
KW haematopoietic disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; gene; ds.
XX
XX Homo sapiens.
XX
XX US2003059908-A1.
XX
XX 27-MAR-2003.
XX
XX 07-MAR-2002; 2002US-00091504.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231422P.
PR 08-SEP-2000; 2000US-0231433P.
PR 08-SEP-2000; 2000US-0231443P.
PR 08-SEP-2000; 2000US-0231444P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234423P.
PR 21-SEP-2000; 2000US-0234423P.
PR 21-SEP-2000; 2000US-0234474P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250316P.
PR 01-DEC-2000; 2000US-0250319P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0256678P.
PR 17-JAN-2001; 2001US-00764869.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-743766/70.
DR
XX
XX
XX New cardiovascular system related polynucleotides and polypeptides,
PT useful for preventing, treating, or ameliorating a medical condition,
PT such as cancer of cardiovascular tissues and cancer metastases.
XX
XX Claim 1; SEQ ID NO 1740; 262pp; English.
XX
XX The invention relates to human cardiovascular system related polypeptides
CC and the polynucleotides encoding them. The polypeptides, polynucleotides
CC and antibodies to the polypeptides are useful for diagnosing a
CC pathological condition or a susceptibility to a pathological condition,
CC for preventing, treating, or ameliorating a medical condition, such as
CC cancer of cardiovascular system tissues, proliferative disorders, foetal
CC and developmental abnormalities, haematopoietic disorders, diseases of
CC the immune system, AIDS, autoimmune diseases (e.g., rheumatoid
CC arthritis), inflammation, allergies, neurological disorders (e.g.,
CC Alzheimer's disease, Parkinson's disease), cognitive disorders,
CC schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis,
CC diabetes, atherosclerosis, cardiovascular disorders, angiogenic
CC disorders, kidney disorders, gastrointestinal disorders, pregnancy-
CC related disorders, endocrine disorders and infections. The nucleic acids
CC are also useful for chromosome identification, radiation hybrid mapping
CC or long-range restriction mapping. The polypeptides and polynucleotides
CC may also be used as food additives or preservatives to increase or
CC decrease storage capabilities, fat content or other nutritional
CC components. This sequence represents human cardiovascular system related
CC genomic DNA of the invention.
XX
XX
SQ Sequence 5628 BP; 1449 A; 1395 C; 1432 G; 1352 T; 0 U; 0 Other;
XX
XX
XX Query Match 76.8%; Score 19.2; DB 10; Length 5628;
Best Local Similarity 87.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 4860 AAAAAAAAAATCCCTATTAGAT 4883
RESULT 39
ADJ08352
ID ADJ08352 standard; DNA; 5628 BP.
XX
XX ADJ08352;
AC
XX
DT 04-NOV-2004 (first entry)
XX
XX Human cardiovascular system associated polypeptide-related DNA SeqID1740.
DE
XX
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KM breast neoplasms; liver neoplasms; cardiovascular disorder;
KM cardiac arrest; cerebrovascular disorder; cerebral ischaemia;
KM angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KM ocular disorder; corneal infection; wound healing;
KM epithelial cell proliferation; skin aging; sunburn;
KM organ transplantation; cell culture; tissue regeneration; chemotaxis;
KM food additive; preservative; cardiovascular system associated antigen;
KM

KM nuclear factor kappaB; NFkappaB; promoter element; human; ds.
XX
XX Homo sapiens.
OS
XX US2004005575-A1.
PN
XX
XX 08-JAN-2004.
PD
XX
XX 26-AUG-2002; 2002US-00227577.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190766P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227099P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228927P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR

PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251888P.
PR 05-DEC-2000; 2000US-0251889P.
PR 06-DEC-2000; 2000US-0251719P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

PR 17-JAN-2001; 2001US-00764869.
PR 07-MAR-2002; 2002US-00091504.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI, 2004-081713/08.
XX
XX
XX New cardiovascular system-related nucleic acid molecule, useful for
PT diagnosing, preventing or treating diseases of the cardiovascular system,
PT and in chromosome mapping, drug screening or in pharmacogenomics.
XX
XX
PS Disclosure; SEQ ID NO 1740; 262bp; English.
XX
XX
CC The invention relates to an isolated nucleic acid molecule encoding a
CC human cardiovascular system associated polypeptide (or antigens), or its
CC fragment. Also included recombinant vectors, recombinant host cells, an
CC isolated human cardiovascular system associated polypeptide (including
CC its fragment, allelic variant, species homologue or epitope), an isolated
CC antibody that binds specifically to a human cardiovascular system
CC associated polypeptide, diagnosing a pathological condition or
CC susceptibility to a pathological condition (comprising determining the
CC presence or absence of a mutation in human cardiovascular system
CC associated nucleic acid and diagnosing a condition based on the presence
CC or absence of the mutation), identifying a binding partner to human
CC cardiovascular system associated polypeptides, the gene corresponding to
CC the human cardiovascular system associated cDNA sequence and identifying
CC an activity in a biological assay comprising expressing the human
CC cardiovascular system associated cDNA in a cell, isolating the
CC supernatant, detecting an activity in a biological assay and identifying
CC the protein in the supernatant having the activity. The human
CC cardiovascular system associated nucleic acids and polypeptides are used
CC to prevent, treat or ameliorate a medical condition (for example in
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep), for
CC example autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders, for example neoplasms of the breast or
CC liver, cardiovascular disorders, for example cardiac arrest,
CC
CC
Query Match 76.8%; Score 19.2; DB 13; Length 5628;
Best Local Similarity 87.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGT 24
DB 4860 AAAAAAAAAATCCTATTAGAT 4883
RESULT 40
ABD32925
ID ABD32925 standard; cDNA, 6134 BP.
XX
XX ABD32925;
AC
XX
XX 18-NOV-2004 (first entry)
DT
XX
XX Human cancer-associated cDNA HR19-002.2.
DE
XX
XX Human; ss; cancer-associated protein; gene; cytosolic; cancer;
KW leukemia; lymphoma; CAP.
XX
XX Homo sapiens.
OS
XX
XX WO2004074320-A2.
PN
XX
XX 02-SEP-2004.
PD
XX
XX 17-FEB-2004; 2004WO-US004730.
PF
XX
XX 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.

PR 15-SEP-2003; 2003US-00663431.
 PR 15-DEC-2003; 2003US-00737318.
 XX
 PA (SAGR-) SAGRES DISCOVERY INC.
 XX
 PI Morris DW, Morris DW, Malandro MS;
 DR WPI; 2004-652914/63.
 DR P-PSDB; ABO84651.
 PT New isolated cancer-associated polynucleotides and polypeptides useful
 PT for diagnosing, preventing or treating cancers, especially lymphoma and
 PT leukemia, or in screening for agents that modulate cancer.
 XX
 PS claim 1; seqid 641; 310pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising at least 10
 CC contiguous nucleotides of any of the 233 polynucleotide sequences given
 CC in the specification, or its complement. The nucleic acids encode cancer-
 CC associated proteins. Also included are an expression vector comprising
 CC the isolated nucleic acid cited above, a host cell comprising the above
 CC recombinant nucleic acid or expression vector, a microarray for detecting
 CC a cancer-associated (CA) nucleic acid comprising at least one probe
 CC comprising at least 10 contiguous nucleotides of any of the above-
 CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
 CC an open reading frame of a CA sequence selected from any of the 95
 CC polynucleotide sequences as mentioned in the specification, or its
 CC complement), an isolated antibody, (or its antigen binding fragment) that
 CC binds to the above polypeptide, a hybridoma that produces the above
 CC monoclonal antibody, a pharmaceutical composition comprising the above
 CC antibody and a pharmaceutical excipient, a kit for detecting cancer
 CC cells comprising the antibody cited above, methods for diagnosing cancer
 CC or for detecting the presence or absence of cancer cells in an
 CC individual, a method for inhibiting a growth of cancer cells in an
 CC individual, a method for delivering a therapeutic agent to cancer cells
 CC in an individual, an electronic library comprising the above
 CC polynucleotide or polypeptide (or their fragments), methods of screening
 CC for anticancer activity or for a bioactive agent capable of modulating
 CC the activity of a CA protein (CAP), methods for detecting cancer
 CC associated with expression of a polypeptide in a test cell sample, a
 CC method for treating cancers and a method for inhibiting the expression of
 CC CA gene in a cell. The composition and methods are useful for detecting,
 CC diagnosing, preventing and treating cancers, especially lymphoma and
 CC leukemia. These may also be used in screening for agents that modulate
 CC cancer. The present sequence is a human CAP cDNA sequence. Note: The
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6134 BP; 1415 A; 1649 C; 1693 G; 1377 T; 0 U; 0 Other;
 Query Match 76.8%; Score 19.2; DB 13; Length 6134;
 Best Local Similarity 87.5%; Pred. No. 5.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
 Db 5960 AAAAAAAAAACTCCCACTCAGAT 5983
 RESULT 41
 ABD32926
 ID ABD32926 standard; cDNA; 6138 BP.
 XX
 AC ABD32926;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human cancer-associated cDNA HR19-002.3.
 XX
 KW Human; sg; cancer-associated protein; gene; cytostatic; cancer;
 KM leukemia; lymphoma; CAP.
 XX

OS Homo sapiens.
 XX
 PN WO2004074320-A2.
 XX
 PD 02-SEP-2004.
 PD
 XX
 PF 17-FEB-2004; 2004MO-US004730.
 PF
 PR 14-FEB-2003; 2003US-00367094.
 PR 14-MAR-2003; 2003US-00388838.
 PR 15-APR-2003; 2003US-00417375.
 PR 13-JUN-2003; 2003US-00461862.
 PR 15-SEP-2003; 2003US-00663431.
 PR 15-DEC-2003; 2003US-00737318.
 XX
 PA (SAGR-) SAGRES DISCOVERY INC.
 XX
 PI Morris DW, Morris DW, Malandro MS;
 DR WPI; 2004-652914/63.
 DR P-PSDB; ABO84652.
 XX
 PT New isolated cancer-associated polynucleotides and polypeptides useful
 PT for diagnosing, preventing or treating cancers, especially lymphoma and
 PT leukemia, or in screening for agents that modulate cancer.
 XX
 PS claim 1; seqid 643; 310pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising at least 10
 CC contiguous nucleotides of any of the 233 polynucleotide sequences given
 CC in the specification, or its complement. The nucleic acids encode cancer-
 CC associated proteins. Also included are an expression vector comprising
 CC the isolated nucleic acid cited above, a host cell comprising the above
 CC recombinant nucleic acid or expression vector, a microarray for detecting
 CC a cancer-associated (CA) nucleic acid comprising at least one probe
 CC comprising at least 10 contiguous nucleotides of any of the above-
 CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
 CC an open reading frame of a CA sequence selected from any of the 95
 CC polynucleotide sequences as mentioned in the specification, or its
 CC complement), an isolated antibody, (or its antigen binding fragment) that
 CC binds to the above polypeptide, a hybridoma that produces the above
 CC monoclonal antibody, a pharmaceutical composition comprising the above
 CC antibody and a pharmaceutical excipient, a kit for detecting cancer
 CC cells comprising the antibody cited above, methods for diagnosing cancer
 CC or for detecting the presence or absence of cancer cells in an
 CC individual, a method for inhibiting a growth of cancer cells in an
 CC individual, a method for delivering a therapeutic agent to cancer cells
 CC in an individual, an electronic library comprising the above
 CC polynucleotide or polypeptide (or their fragments), methods of screening
 CC for anticancer activity or for a bioactive agent capable of modulating
 CC the activity of a CA protein (CAP), methods for detecting cancer
 CC associated with expression of a polypeptide in a test cell sample, a
 CC method for treating cancers and a method for inhibiting the expression of
 CC CA gene in a cell. The composition and methods are useful for detecting,
 CC diagnosing, preventing and treating cancers, especially lymphoma and
 CC leukemia. These may also be used in screening for agents that modulate
 CC cancer. The present sequence is a human CAP cDNA sequence. Note: The
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6138 BP; 1416 A; 1649 C; 1695 G; 1378 T; 0 U; 0 Other;
 Query Match 76.8%; Score 19.2; DB 13; Length 6138;
 Best Local Similarity 87.5%; Pred. No. 5.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
 Db 5964 AAAAAAAAAACTCCCACTCAGAT 5987
 RESULT 42

ABL2680	ID ABL2680 standard; DNA; 31271 BP.
XX AC	ABL2680;
XX DT	26-MAR-2002 (first entry)
DE XX	Drosophila melanogaster genomic polynucleotide SEQ ID NO 31513.
KW XX	Drosophila; developmental biology; cell signalling; insecticide;
OS XX	pharmaceutical; gene; ds.
PN OS	Drosophila melanogaster.
PD PN	WO200171042-A2.
PF PD	27-SEP-2001.
PR PF	23-MAR-2001; 2001WO-US009231.
PT PR	23-MAR-2000; 2000US-0191637P.
PS PT	11-JUL-2000; 2000US-00614150.
PP PA	(PEKE) PE CORP NY.
PI PP	Venter JC, Adams M, Li FWD, Myers EW;
DR PI	WPI; 2001-656860/75.
DT DR	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
CC PS	Claim 1; SEQ ID NO 31513; 21bp + Sequence Listing; English.
CC CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABU30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SO CC	Sequence 31271 BP; 8794 A; 6300 C; 6590 G; 9587 T; 0 U; 0 Other;
Query Match	76.8%; Score 19.2; DB 4; Length 31271;
Best Local Similarity	87.5%; Pred.No.5.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0	
QY	1 AAAAAAAAAAGTCCCATTCAATG 24 DB 6323 AAAAAAATAGTTCATTAAAGAT 6346
RESULT 43	
ADNO1773/c	
ID ADNO1773	standard; DNA; 65454 BP.
AC ADNO1773;	
DT	29-JUL-2004 (first entry)
DX Human huntingtin interacting protein 1, HIP1, genomic sequence.	
XX Human; antisense; ds; Huntingtin interacting protein 1; HIP1;	
KW cellular apoptosis; Huntington's disease; gene; chromosome 7q11.23.	
OS Homo sapiens.	
PN US2004092465-A1.	

```

XX 13-MAY-2004.
XX
XX
XX 11-NOV-2002; 2002US-00293864.
XX
XX 11-NOV-2002; 2002US-00293864.
XX
XX 11-NOV-2002; 2002US-00293864.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX PA
XX
XX Dobie KW;
XX
XX WPI, 2004-374983/35.
XX
XX
XX New compound that modulates huntingtin interacting protein 1 expression,
XX useful in treating an animal having a disease or condition involving
XX dysregulation of cellular apoptosis.
XX
XX
XX Example 15; SEQ ID NO 11; 85pp; English.
XX
XX The invention relates to a compound targeted to a nucleic acid molecule
XX encoding huntingtin interacting protein 1, H1P1. The compound, 8-80
XX nucleobases in length, is an antisense oligonucleotide, where the
XX compound specifically hybridises with the nucleic acid molecule encoding
XX huntingtin interacting protein 1 comprising a sequence appearing as
XX ADN01766 and inhibits the expression of huntingtin interacting
XX protein 1 in cells or tissues, screening for a modulator of huntingtin
XX interacting protein 1, a diagnostic method for identifying a disease
XX state, a kit or assay device comprising the compound and treating an
XX animal having a disease or condition associated with huntingtin
XX interacting protein 1 compound so that expression of huntingtin
XX interacting protein 1 is inhibited. The compound and the methods are
XX useful in treating an animal having a disease or condition involving
XX dysregulation of cellular apoptosis e.g. Huntington's disease. The H1P1
XX gene is located on chromosome 7q41.23. The present sequence is the H1P1
XX genomic DNA, a target for the antisense oligonucleotides.
XX
XX
XX Sequence 65454 BP; 16579 A; 15844 C; 16495 G; 16536 T; 0 U; 0 Other;
XX
XX
XX Query Match 76.8%; Score 19.2; DB 12; Length 65454;
XX
XX Best Local Similarity 87.5%; Pred. No. 6e+02;
XX
XX Match 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX
XX 1 AAAAAAAAAAGTCCCAATTCAGAT 24
XX |||||
XX Db 2046 AAAAAAAAAAGTCCCACTCGCT 2023
XX
XX
XX RESULT 44
XX ADO97351/C
XX ID ADO97351 standard; DNA; 66315 BP.
XX
XX ADO97351;
XX
XX 07-OCT-2004 (first entry)
XX
XX Human cancer associated sequence HD08-036, SEQ ID 328.
XX
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
XX Homo sapiens.
XX
XX WO2004060304-A2.
XX
XX 22-JUN-2004.
XX
XX 22-DEC-2003; 2003WO-US041389.
XX
XX 27-DEC-2002; 2002US-00330773.
XX
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX

```

```

XX DR WPI: 2004-543781/52.
XX
XX PT New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX
XX PS Claim 1; SEQ ID NO 328; 199bp; English.
XX
XX CC The present invention relates to cancer associated sequences (ADQ97025-
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 66315 BP; 18250 A; 13771 C; 13088 G; 19203 T; 0 U; 2003 Other;

Query Match 76.8%; Score 19.2; DB 12; Length 66315;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
DB 39105 AAAAAAAAAAGTCCCAACAGAT 39082

RESULT 45
ABD32923_7
Continuation (8 of 8) of ABD32923 from base 700001 (Human cancer-associated genomic DNA
WP Sequence split into 8 fragments LOCUS ABD32923 Accession ABD32923
WP Fragment Name Begin End
WP ABD32923_0 1 110000
WP ABD32923_1 100001 210000
WP ABD32923_2 200001 310000
WP ABD32923_3 300001 410000
WP ABD32923_4 400001 510000
WP ABD32923_5 500001 610000
WP ABD32923_6 600001 710000
WP ABD32923_7 700001 788759

Query Match 76.8%; Score 19.2; DB 13; Length 88759;
Best Local Similarity 87.5%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
DB 78585 AAAAAAAAAAGTCCCACTGAT 78608

RESULT 46
ADP81772/c
ID ADP81772 standard; DNA; 67674 BP.
XX
XX AC ADP81772;
XX
XX DT 26-AUG-2004 (first entry)
XX
XX DE Human MD-1 RP105-associated complementary DNA.
XX
XX KM MD-1 RP105-associated; MD-1; MD1; autoimmune disorder; gene therapy;
XX KM human; gene; ds.
XX
XX OS Homo sapiens.
XX
XX PN US2004110146-A1.
XX
XX PD 10-JUN-2004.
XX
XX PF 09-DEC-2002; 2002US-00316242.
XX
XX PR 09-DEC-2002; 2002US-00316242.
XX
XX PT

```

```

PA (ISIS-) ISIS PHARM INC.
XX
XX PI Dobie KW;
XX
XX DR WPI: 2004-440335/41.
XX DR REFSEQ; NT_023412.8.
XX
XX PT New oligonucleotide compound that inhibits expression of MD-1 RP105-
XX PT associated; useful for preparing a composition for treating autoimmune
XX PT disorder.
XX
XX PS Example 15; SEQ ID NO 11; 63bp; English.
XX
XX CC The invention relates to compounds, compositions and methods for
XX CC modulating the expression of MD-1 RP105-associated (also called as MD-1
XX CC and MD1) DNA. The composition comprise antisense oligonucleotides
XX CC targeted to MD-1 RP105-associated DNA. The compound is useful for
XX CC preparing a composition for treating autoimmune disorder. It is also
XX CC useful in gene therapy. The present sequence is human MD-1 RP105-
XX CC associated complementary DNA. This sequence is used to illustrate the
XX CC method of the invention.
XX
XX SQ Sequence 67674 BP; 19614 A; 15528 C; 14518 G; 18014 T; 0 U; 0 Other;

Query Match 76.0%; Score 19; DB 12; Length 67674;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAAT 19
DB 40339 AAAAAAAAAAGTCCCAAT 40321

RESULT 47
ADP09628/c
ID ADP09628 standard; DNA; 2971 BP.
XX
XX AC ADP09628;
XX
XX DT 26-AUG-2004 (first entry)
XX
XX DE Rice alternative oxidase genomic DNA without the T-DNA insert Segid 19.
XX
XX KM rice; AOX1; GUS; insertional mutagenesis; beta-glucuronidase;
XX KM biotic stress resistance; pesticide; herbicide; plant; ds;
XX KM hygromycin phosphotransferase; HGH; T-DNA; transfer-DNA;
XX KM alternative oxidase.
XX
XX OS Oryza sativa.
XX
XX FH Key Location/Qualifiers
XX FT misc_feature 577..578
XX FT /tag= a
XX FT /note= "Location of T-DNA insert in rice line IB-164-43"
XX
XX PN WO2004046357-A1.
XX
XX PD 03-JUN-2004.
XX
XX PF 14-NOV-2003; 2003WO-KR002461.
XX
XX PR 15-NOV-2002; 2002US-0427166P.
XX
XX PA (POSC-) POSCO.
XX PA (POST-) POSTECH FOUND.
XX
XX PI An G, Ryu C, Han J, Kang H, An K;
XX
XX DR WPI: 2004-449751/42.
XX DR P-PSDB; ADP09662.
XX
XX PT New organ preferential nucleic acids and polypeptides, useful in
XX PT producing rice plants with desired characteristics and which are

```

PT resistant to herbicide, plant pathogen, fungi, bacteria, virus, insect,
PT nematode, and stress.
XX
PS Claim 1; SEQ ID NO 19; 236bp; English.
XX
CC This invention relates to a novel method of identifying nucleic acid
CC molecules, or fragments thereof, that are expressed in an organ
CC preferential manner in rice. Specifically, it refers to producing rice
CC cell lines that carry tagged genes modified by T-DNA/GUS based
CC insertional mutagenesis, where the GUS portion of the insert is the
CC promoterless beta-glucuronidase reporter gene that can only be expressed
CC when it is inserted into an active gene. The present invention describes
CC generating transformed rice lines containing transfer-DNA (T-DNA)
CC insertions that also carry a gene encoding the selectable marker
CC hygromycin phosphotransferase (HPT). Accordingly, this method can be used
CC to produce rice plants with desirable characteristics including increased
CC grain yield and nutritional content, resistance to biotic stress,
CC pesticides, herbicides or insects, altered morphology or improved
CC characteristics referring to shape, taste or cooking quality of the
CC grain. This polynucleotide sequence represents the genomic DNA of a rice
CC gene that will be T-DNA/GUS tagged for expression analyses, given in an
CC exemplification of the invention.
XX
SQ Sequence 2971 BP; 665 A; 735 C; 749 G; 822 T; 0 U; 0 Other;
Query Match 75.2%; Score 18.8; DB 12; Length 2971;
Best Local Similarity 90.9%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 4 AAAAAAGTCCCAATTCAGATA 25
Db 2490 AAAAAACGCGCCCAATTCAGATA 2469
RESULT 48
AAC90666/c
ID AAC90666 standard; DNA; 3696 BP.
XX
AC AAC90666;
XX
DT 20-MAR-2001 (first entry)
XX
DE Strawberry flowering regulation protein coding sequence SEQ ID NO: 37.
XX
KM Strawberry; flowering regulation; floral homeotic gene; field crop;
XX harvesting; fruit production; ds.
OS Fragaria vesca.
XX
PN WO200071722-A1.
XX
PD 30-NOV-2000.
XX
PE 24-MAY-2000; 2000WO-US014297.
XX
PR 25-MAY-1999; 99US-00318789.
PR 24-MAY-2000; 2000US-00318789.
XX
PA (DNAP) DNA PLANT TECHNOLOGY CORP.
XX
PI Oeller P, Guttersen N;
XX
DR WPI; 2001-025165/03.
DR P-PSDB; AAB50269.
XX
PT Novel nucleic acid involved in controlling plant flowering processes is
PT useful for generating transgenic plants, in particular strawberry plants
PT having altered flowering behavior such as early, delayed or day-neutral
PT flowering.
XX
PS Disclosure; Page 82-87; 97tp; English.
XX The present invention provides the nucleic acid and protein sequences of

CC a number of proteins from the strawberry which are involved in the
CC regulation of flowering. These were identified using primers based on the
CC homologous sequences from A. thaliana, B. napus and R. sativus. They can
CC be used in the production of transgenic field crops whose flowering is
CC regulated and the time of fruiting and harvesting can be manipulated
XX
SQ Sequence 3696 BP; 1175 A; 618 C; 648 G; 1255 T; 0 U; 0 Other;
Query Match 75.2%; Score 18.8; DB 4; Length 3696;
Best Local Similarity 90.9%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 4 AAAAAAGTCCCAATTCAGATA 25
Db 2926 AAAAAAGTCCCAATTCAGATA 2905
RESULT 49
AAK69744
ID AAK69744 standard; DNA; 5642 BP.
XX
AC AAK69744;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24556.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229286P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232088P.
PR 08-SEP-2000; 2000US-0232089P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239937P.
PR 13-OCT-2000; 2000US-0239938P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.

PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251865P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254057P.
PR 05-JAN-2001; 2001US-0259678P.

PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
PT
XX
XX Disclosure; SEQ ID NO 24556; 3071bp + Sequence Listing; English.
PS
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
XX Sequence 5642 BP; 1411 A; 1255 C; 1549 G; 1427 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 4; Length 5642;
Best Local Similarity 90.9%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGCCCAATCAG 22
Db 1298 AAAAAAAAAAGCTATATTCAG 1319

RESULT 50
AAS34694
ID AAS34694 standard; DNA; 5642 BP.
XX

AC AAS34694;
XX 17-DEC-2001 (first entry)
XX Human DNA for a novel foetal antigen, SEQ ID No 2118.
DE
XX Human; foetal tissue antigen; ds; antiinflammatory; neuroprotective;
immunomodulator; cardiovascular; cytostatic; nephrotoxic;
KM cardiovascular; autoimmune disease; rheumatoid arthritis;
KM hyperproliferative disorder; breast neoplasm; cancer;
KM cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KM cerebral ischaemia; angiogenesis; nervous system disorder;
KM Alzheimer's disease; infection; ocular disorder; corneal infection;
KM wound healing; epithelial cell proliferation; food additive.
XX Homo sapiens.
XX WO20015312-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001321.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 22-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235844P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236378P.
PR 29-SEP-2000; 2000US-0236379P.
PR 29-SEP-2000; 2000US-0236380P.
PR 29-SEP-2000; 2000US-0236381P.
PR 29-SEP-2000; 2000US-0236382P.
PR 29-SEP-2000; 2000US-0236383P.
PR 29-SEP-2000; 2000US-0236384P.
PR 29-SEP-2000; 2000US-0236385P.
PR 29-SEP-2000; 2000US-0236386P.
PR 29-SEP-2000; 2000US-0236387P.
PR 29-SEP-2000; 2000US-0236388P.
PR 29-SEP-2000; 2000US-0236389P.
PR 29-SEP-2000; 2000US-0236390P.
PR 29-SEP-2000; 2000US-0236391P.
PR 29-SEP-2000; 2000US-0236392P.
PR 29-SEP-2000; 2000US-0236393P.
PR 29-SEP-2000; 2000US-0236394P.
PR 29-SEP-2000; 2000US-0236395P.
PR 29-SEP-2000; 2000US-0236396P.
PR 29-SEP-2000; 2000US-0236397P.
PR 29-SEP-2000; 2000US-0236398P.
PR 29-SEP-2000; 2000US-0236399P.
PR 29-SEP-2000; 2000US-0236400P.
PR 29-SEP-2000; 2000US-0236401P.
PR 29-SEP-2000; 2000US-0236402P.
PR 29-SEP-2000; 2000US-0236403P.
PR 29-SEP-2000; 2000US-0236404P.
PR 29-SEP-2000; 2000US-0236405P.
PR 29-SEP-2000; 2000US-0236406P.
PR 29-SEP-2000; 2000US-0236407P.
PR 29-SEP-2000; 2000US-0236408P.
PR 29-SEP-2000; 2000US-0236409P.
PR 29-SEP-2000; 2000US-0236410P.
PR 29-SEP-2000; 2000US-0236411P.
PR 29-SEP-2000; 2000US-0236412P.
PR 29-SEP-2000; 2000US-0236413P.
PR 29-SEP-2000; 2000US-0236414P.
PR 29-SEP-2000; 2000US-0236415P.
PR 29-SEP-2000; 2000US-0236416P.
PR 29-SEP-2000; 2000US-0236417P.
PR 29-SEP-2000; 2000US-0236418P.
PR 29-SEP-2000; 2000US-0236419P.
PR 29-SEP-2000; 2000US-0236420P.
PR 29-SEP-2000; 2000US-0236421P.
PR 29-SEP-2000; 2000US-0236422P.
PR 29-SEP-2000; 2000US-0236423P.
PR 29-SEP-2000; 2000US-0236424P.
PR 29-SEP-2000; 2000US-0236425P.
PR 29-SEP-2000; 2000US-0236426P.
PR 29-SEP-2000; 2000US-0236427P.
PR 29-SEP-2000; 2000US-0236428P.
PR 29-SEP-2000; 2000US-0236429P.
PR 29-SEP-2000; 2000US-0236430P.
PR 29-SEP-2000; 2000US-0236431P.
PR 29-SEP-2000; 2000US-0236432P.
PR 29-SEP-2000; 2000US-0236433P.
PR 29-SEP-2000; 2000US-0236434P.
PR 29-SEP-2000; 2000US-0236435P.
PR 29-SEP-2000; 2000US-0236436P.
PR 29-SEP-2000; 2000US-0236437P.
PR 29-SEP-2000; 2000US-0236438P.
PR 29-SEP-2000; 2000US-0236439P.
PR 29-SEP-2000; 2000US-0236440P.
PR 29-SEP-2000; 2000US-0236441P.
PR 29-SEP-2000; 2000US-0236442P.
PR 29-SEP-2000; 2000US-0236443P.
PR 29-SEP-2000; 2000US-0236444P.
PR 29-SEP-2000; 2000US-0236445P.
PR 29-SEP-2000; 2000US-0236446P.
PR 29-SEP-2000; 2000US-0236447P.
PR 29-SEP-2000; 2000US-0236448P.
PR 29-SEP-2000; 2000US-0236449P.
PR 29-SEP-2000; 2000US-0236450P.
PR 29-SEP-2000; 2000US-0236451P.
PR 29-SEP-2000; 2000US-0236452P.
PR 29-SEP-2000; 2000US-0236453P.
PR 29-SEP-2000; 2000US-0236454P.
PR 29-SEP-2000; 2000US-0236455P.
PR 29-SEP-2000; 2000US-0236456P.
PR 29-SEP-2000; 2000US-0236457P.
PR 29-SEP-2000; 2000US-0236458P.
PR 29-SEP-2000; 2000US-0236459P.
PR 29-SEP-2000; 2000US-0236460P.
PR 29-SEP-2000; 2000US-0236461P.
PR 29-SEP-2000; 2000US-0236462P.
PR 29-SEP-2000; 2000US-0236463P.
PR 29-SEP-2000; 2000US-0236464P.
PR 29-SEP-2000; 2000US-0236465P.
PR 29-SEP-2000; 2000US-0236466P.
PR 29-SEP-2000; 2000US-0236467P.
PR 29-SEP-2000; 2000US-0236468P.
PR 29-SEP-2000; 2000US-0236469P.
PR 29-SEP-2000; 2000US-0236470P.
PR 29-SEP-2000; 2000US-0236471P.
PR 29-SEP-2000; 2000US-0236472P.
PR 29-SEP-2000; 2000US-0236473P.
PR 29-SEP-2000; 2000US-0236474P.
PR 29-SEP-2000; 2000US-0236475P.
PR 29-SEP-2000; 2000US-0236476P.
PR 29-SEP-2000; 2000US-0236477P.
PR 29-SEP-2000; 2000US-0236478P.
PR 29-SEP-2000; 2000US-0236479P.
PR 29-SEP-2000; 2000US-0236480P.
PR 29-SEP-2000; 2000US-0236481P.
PR 29-SEP-2000; 2000US-0236482P.
PR 29-SEP-2000; 2000US-0236483P.
PR 29-SEP-2000; 2000US-0236484P.
PR 29-SEP-2000; 2000US-0236485P.
PR 29-SEP-2000; 2000US-0236486P.
PR 29-SEP-2000; 2000US-0236487P.
PR 29-SEP-2000; 2000US-0236488P.
PR 29-SEP-2000; 2000US-0236489P.
PR 29-SEP-2000; 2000US-0236490P.
PR 29-SEP-2000; 2000US-0236491P.
PR 29-SEP-2000; 2000US-0236492P.
PR 29-SEP-2000; 2000US-0236493P.
PR 29-SEP-2000; 2000US-0236494P.
PR 29-SEP-2000; 2000US-0236495P.
PR 29-SEP-2000; 2000US-0236496P.
PR 29-SEP-2000; 2000US-0236497P.
PR 29-SEP-2000; 2000US-0236498P.
PR 29-SEP-2000; 2000US-0236499P.
PR 29-SEP-2000; 2000US-0236500P.


```

PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251855P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0255678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX DR WPI; 2001-488782/53.
XX
XX FX New polynucleotides and polypeptides for diagnosing, treating, preventing
PT or prognosing e.g. diseases or disorders of the nervous, musculoskeletal,
PT excretory, gastrointestinal, reproductive, and respiratory systems.
XX
XX PS Disclosure; SEQ ID NO 2118; 642pp; English.
XX
XX CC The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the antigens can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities, fat content, lipid, protein, carbohydrate,
CC vitamins, minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence is a
CC genomic DNA fragment from a gene encoding a foetal antigen of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly

```

```

Query Match 75.2%; Score 18.8; DB 5; Length 5642;
Best Local Similarity 90.9%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1 AAAAAAAAAAGTCCATTTCAG 22
Db 1298 AAAAAAAAAAGTCAATATTCAG 1319

```

```

Search completed: December 14, 2005, 02:43:04
Job time : 210.2 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:34:03 ; Search time 1757.1 Seconds
(without alignments)
667.586 Million cell updates/sec

Title: US-10-681-773-7
Perfect score: 25
Sequence: 1 aaaaaaaaaagtcacatcagata 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 segs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

EST:*

1: gb_est1:.*
2: gb_est2:.*
3: gb_est3:.*
4: gb_est4:.*
5: gb_est5:.*
6: gb_est6:.*
7: gb_est7:.*
8: gb_est8:.*
9: gb_est9:.*
10: gb_est10:.*
11: gb_est11:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	93.6	776	9	AQ870656 nbe00400
2	22.4	89.6	553	9	AZ434046 1M0220E18
3	21.8	87.2	533	10	CNS00PBY
4	21.8	87.2	590	10	CM139468
5	21.8	87.2	719	10	CM091591
6	21.8	87.2	777	8	CX451695
7	21.8	87.2	997	9	BI465187
8	21.8	87.2	1000	9	B08139
9	21.8	87.2	1003	9	BI2235
10	21.8	87.2	1030	9	BI2507
11	21.4	85.6	331	3	BI627421
12	21.4	85.6	659	3	BP704125
13	21.4	85.6	761	10	AG557295
14	20.8	83.2	182	7	CM640041
15	20.8	83.2	214	1	AL785515
16	20.8	83.2	350	3	BI615833
17	20.8	83.2	375	2	BE101746
18	20.8	83.2	492	8	CX329379
19	20.8	83.2	514	6	CF112690
20	20.8	83.2	514	8	CF112690
21	20.8	83.2	523	1	AL969417
22	20.8	83.2	529	3	BP709610

23	20.8	83.2	551	8	CX474359
24	20.8	83.2	560	7	CM949747
25	20.8	83.2	567	8	CX418273
26	20.8	83.2	567	8	CX450676
27	20.8	83.2	578	8	CX417651
28	20.8	83.2	589	7	CR429411
29	20.8	83.2	614	8	CX496283
30	20.8	83.2	655	8	CX413722
31	20.8	83.2	655	8	DR902299
32	20.8	83.2	664	5	BU418890
33	20.8	83.2	698	7	CR560919
34	20.8	83.2	710	7	CR560918
35	20.8	83.2	711	7	CR413014
36	20.8	83.2	715	8	CX401299
37	20.8	83.2	721	8	CR565819
38	20.8	83.2	734	8	CX916844
39	20.8	83.2	757	5	BR727946
40	20.8	83.2	757	8	CX359920
41	20.8	83.2	778	8	CX387336
42	20.8	83.2	793	8	CX391469
43	20.8	83.2	795	5	BR738214
44	20.8	83.2	802	8	CX329053
45	20.8	83.2	813	8	CX362328
46	20.8	83.2	815	7	CR569461
47	20.8	83.2	821	7	CR413015
48	20.8	83.2	824	8	CX329054
49	20.8	83.2	852	10	CL101305
50	20.8	83.2	851	7	CK315320
51	20.4	81.6	264	10	CM905537
52	20.4	81.6	488	10	CZ415534
53	20.4	81.6	559	9	AQ356837
54	20.4	81.6	605	9	AQ477054
55	20.4	81.6	684	10	CG854333
56	20.4	81.6	703	6	CD367099
57	20.4	81.6	703	9	BR949068
58	20.4	81.6	749	10	AG381491
59	20.4	81.6	750	9	BZ058921
60	20.2	80.8	109	3	BP702142
61	20.2	80.8	148	1	A1228545
62	20.2	80.8	148	3	BP695277
63	20.2	80.8	154	3	BP675467
64	20.2	80.8	181	3	BP695882
65	20.2	80.8	166	7	CV583054
66	20.2	80.8	209	10	AL946261
67	20.2	80.8	220	9	AZ760441
68	20.2	80.8	230	1	AV257118
69	20.2	80.8	234	3	BP712930
70	20.2	80.8	245	8	CX427015
71	20.2	80.8	260	9	AQ342305
72	20.2	80.8	260	1	AM200277
73	20.2	80.8	271	3	BI675435
74	20.2	80.8	294	1	AV020455
75	20.2	80.8	299	1	A1449377
76	20.2	80.8	349	1	AA817980
77	20.2	80.8	358	2	BE506958
78	20.2	80.8	400	1	AV801410
79	20.2	80.8	402	2	BF286272
80	20.2	80.8	411	7	CV627249
81	20.2	80.8	417	2	BF615912
82	20.2	80.8	421	9	AZ300408
83	20.2	80.8	422	1	BE192215
84	20.2	80.8	426	1	A1133803
85	20.2	80.8	461	7	CN108169
86	20.2	80.8	468	9	AZ035656
87	20.2	80.8	479	8	CX424405
88	20.2	80.8	485	9	AZ121992
89	20.2	80.8	485	9	AZ014314
90	20.2	80.8	494	7	CR629336
91	20.2	80.8	494	7	CR629336
92	20.2	80.8	500	11	CR075004
93	20.2	80.8	504	3	BM165368
94	20.2	80.8	511	3	BM167726
95	20.2	80.8			

CX474359	JGI_XZG49
CM949747	XXXB155_N
CX418273	JGI_XZG65
CX450676	JGI_XZG22
CX417651	JGI_XZG65
CR429411	CR429411
CX496283	JGI_XZG38
CX413722	JGI_XZT49
DR902299	JGI_XZT61
BU418890	603959342
CR560919	CR560919
CR560918	CR560918
CR413014	CR413014
CX401299	JGI_XZT49
CR565819	CR565819
CX916844	JGI_CAN5
BR727946	BR727946
CX359920	JGI_XZT95
CX387336	JGI_XZT21
CX391469	JGI_XZT38
BR738214	BR738214
CX329053	JGI_XZT67
CX362328	JGI_XZT42
CR569461	CR569461
CR413015	CR413015
CX329054	JGI_XZT67
CL101305	ISB1-3BE1
CK315320	SB02024B1
CM905537	RPIC142_14
CZ415534	1011237_R
AQ356837	CITBI-B1-
AQ477054	CITBI-B1-
CG854333	ZMMBB035
CD367099	UI-H-FP2-
BR949068	odh71905-
AG381491	Mus_muscu
BZ058921	11632E03.
BP702142	BP702142
A1228545	EST725240
BP695277	BP695277
BP675467	BP675467
BP695882	BP695882
CV583054	L_AK-aa0
AL946261	Arabidops
AZ760441	1M0554N12
AV257118	AV257118
BP712930	BP712930
CX427015	JGI_XZG13
AQ342305	RPCT11-11
AM200277	da17h08.Y
BI675435	daa97410.
AV020455	AV020455
A1449377	wt10d03.x
AA817980	UI-R-A0-a
BE506958	db89a04.Y
BP684767	BP684767
AV801410	AV801410
BF286272	EST450863
CV627249	Mdrt60110
BF615912	de20a05.Y
AZ300408	RPCT-23-1
BE192215	db89a04.x
A1133803	BSBMFSZ4
CN108169	ECSCA31B
AZ035656	RPCT-23-3
CX424405	JGI_XZG64
AZ121992	RPCT-23-1
AZ014314	RPCT-23-3
CG991942	CH240_151
CR629336	DRZP468J
CR075004	Forward_B
BM165368	EST567891
BM167726	EST570249

C 96	20.2	80.8	512	3	BE678824	BE678824
C 97	20.2	80.8	514	9	A2467699	IM0279H14
98	20.2	80.8	515	9	B5081.5	CIT978SK-10
99	20.2	80.8	524	7	CO420150	GGRHC100
100	20.2	80.8	535	10	CO093420	GR_Ba14P
101	20.2	80.8	536	10	AG225601	Lotus
102	20.2	80.8	558	7	CN117165	EC2CAA46A
C 103	20.2	80.8	560	10	BX994105	Reverse s
C 104	20.2	80.8	575	5	BQ386868	NISC_mn20
C 105	20.2	80.8	575	10	CE497075	t1gr-gss-
C 106	20.2	80.8	589	7	CN099847	EC2CAA16D
C 107	20.2	80.8	603	7	CN099846	EC2CAA16D
C 108	20.2	80.8	603	8	CX346639	JGT-XZT43
C 109	20.2	80.8	606	9	A2842709	2M0141118
C 110	20.2	80.8	631	6	CA988492	AGENCOURT
C 111	20.2	80.8	634	7	CN108170	EC2CAA31B
C 112	20.2	80.8	637	10	CE503284	t1gr-gss-
C 113	20.2	80.8	650	6	CD848378	DHAC0052
C 114	20.2	80.8	654	8	CX477767	JGT-XZG20
C 115	20.2	80.8	660	9	CE232282	t1gr-gss-
C 116	20.2	80.8	666	7	CN095710	EC2CAA10B
C 117	20.2	80.8	671	9	A2269879	RPCI-23-1
C 118	20.2	80.8	674	7	CN109563	EC2CAA33D
C 119	20.2	80.8	683	3	BQ118778	EST604354
C 120	20.2	80.8	687	11	CR270022	Forward s
C 121	20.2	80.8	691	8	DR843050	JGT-CABE9
C 122	20.2	80.8	706	8	DR436304	ACAB-aaa4
C 123	20.2	80.8	717	8	CX428139	JGT-XZG18
C 124	20.2	80.8	721	10	AG409282	Mus_muscu
C 125	20.2	80.8	725	6	CB008651	VVC058H04
C 126	20.2	80.8	735	10	CZ407435	1003138 R
C 127	20.2	80.8	746	9	CC913903	t080m04ba
C 128	20.2	80.8	756	10	AG486212	Mus_muscu
C 129	20.2	80.8	764	5	BX689407	BX689407
C 130	20.2	80.8	769	8	DR723603	AGENCOURT
C 131	20.2	80.8	771	5	BX699349	BX699349
C 132	20.2	80.8	777	8	CX447066	JGT-XZG22
C 133	20.2	80.8	779	8	CX739913	JGT-XZG23
C 134	20.2	80.8	790	7	CN113562	EC2CAA3CD
C 135	20.2	80.8	794	7	CR851771	CR851771
C 136	20.2	80.8	809	8	CX427016	JGT-XZG13
C 137	20.2	80.8	815	5	BX700099	BX700099
C 138	20.2	80.8	817	8	CX454235	JGT-XZG56
C 139	20.2	80.8	822	8	CX739914	JGT-XZG23
C 140	20.2	80.8	829	9	B2139649	CH230-511
C 141	20.2	80.8	833	6	CB561618	AGENCOURT
C 142	20.2	80.8	847	5	BX761936	BX761936
C 143	20.2	80.8	863	5	BH315355	603487670
C 144	20.2	80.8	873	5	BX698012	BX698012
C 145	20.2	80.8	877	6	CD255023	AGENCOURT
C 146	20.2	80.8	880	11	CR207177	Forward s
C 147	20.2	80.8	883	10	AL216145	Tetradon
C 148	20.2	80.8	905	5	BX758691	BX758691
C 149	20.2	80.8	911	5	BQ733745	AGENCOURT
C 150	20.2	80.8	935	7	CO073496	GR_Ea33C

ALIGNMENTS

RESULT 1
A0870656 776 bp DNA linear GSS 03-NOV-1999
LOCUS nbdb0040017r CUGI Rice BAC Library (ECORI) Oryza sativa (japonica
DEFINITION cultivar-group) genomic clone nbdb0040017r, genomic survey
sequence.
ACCESSION A0870656
VERSION A0870656.1 GI:6221107
KEYWORDS GSS.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE

1 (bases 1 to 776)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAACAGCATGACCATG
Class: BAC ends
High quality sequence start: 39
High quality sequence stop: 519.

FEATURES

source

1..776
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbdb0040017r"
/tissue="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACindigo; Site 1: EcoRI; Site 2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa.
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9%. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Query Match 93.6%; Score 23.4; DB 9; Length 776;
Best Local Similarity 96.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Dn 1 AAAAAAAAAAGTCCCAATTGAGAT 25
80 AAAAAAAAAAGTCCCAATTCAGAA 104

RESULT 2
A2434046/c 553 bp DNA linear GSS 03-OCT-2000
LOCUS IM0220E18F Mouse 10kb plasmid UUCGIM library Mus musculus genomic
DEFINITION clone UUCGIM0220E18 F, genomic survey sequence.
ACCESSION A2434046
VERSION A2434046.1 GI:10558059
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 553)

TITLE	JOURNAL COMMENT		
TITLE	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duvall, B., Hamill, C., Jellam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R. Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts Unpublished (2000)		
CONTACT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: dduuu@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0220 Row: E Column: 18 Seq primer: CGTGTGTAACGACGCGCCAGT Class: plasmid ends High quality sequence stop: 553.		
FEATURES	Location/Qualifiers 1..553 /organism="Mus musculus" /mol_type="genomic DNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="U06C1M0220E18" /sex="Male" /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-" /clone_lib="Mouse 10kb plasmid U06C1M library" /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114 gb AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."		
ORIGIN	Query Match Best Local Similarity 89.6%; Score 22.4; DB 9; Length 553; Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0; QY 1 AAAAAAAAAAGTCCCAATTTCAGAT 24 Db 325 AAAAAAAAAAGTCCCAATTTCAGAT 302 		
RESULT 3	CNS00PBV 533 bp DNA linear GSS 28-JUN-1999 CNS00PBV LOCUS DEFINITION ICG library from strain Columbia of Arabidopsis thaliana, genomic survey sequence. ACCESSION AL084313.1 GI:5285453 VERSION KEYWORDS SOURCE ORGANISM Arabidopsis thaliana (thale cress) Arabidopsis thaliana Arabidopsis thaliana Bakariotae: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;		

REFERENCE	Authors	Journal	Title	Journal
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 533)	Salanoubat, M., Choisne, N., Artiguenave, F., Brotlier, P., Wincker, P., Samsen, D., Saurin, W., Weissenbach, J. and Quetier, F.	Unpublished	2 (bases 1 to 533)	Genoscope.
REFERENCE	Authors	Journal	Title	Journal
Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	- Web : www.genoscope.cns.fr)	Location/Qualifiers	1. 533	/organism="Arabidopsis thaliana"
FEATURES	source			
Query Match	87.2%;	Score 21.8;	DB 10;	Length 533;
Best Local Similarity	92.0%;	Pred. No. 6.9e+02;		
Matches	23;	Conservative	0;	Mismatches 2;
Indels	0;	Gaps	0;	
Qy	1	AAAAAAAAAAGTCCCAATTCAGATA	25	
Db	253	AAAAAAAAAAGTCCCAATTCAGAAA	277	
RESULT 4				
LOCUS	CM139468/C	590 bp	DNA	linear
DEFINITION	104 529 11135254 116 34905 085 Sorghum methylation filtered library (LibID: 104) Sorghum bicolor genomic clone 11135254, genomic survey			
ACCESSION	CM139468			
VERSION	CM139468			
KEYWORDS	CM139468.1	GI:54832015		
SOURCE	Sorghum bicolor (Sorghum)			
ORGANISM	Sorghum bicolor			
REFERENCE	Authors	Journal	Title	Journal
Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 590)	Bedell, J. A., Budiman, M. A., Nunberg, A., Citek, R. W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., Rabbinowitz, P. D., Lakey, N., McCombie, W. R., Jeddeloh, J. A. and Mortlensen, R. A.	Sorghum genome sequencing by methylation filtration	15660154	Contact: Bedell JA
COMMENT	4041 Forest Park Ave, St. Louis, MO 63108, USA			
TITLE	Orion Genomics, LLC			
JOURNAL	4041 Forest Park Ave, St. Louis, MO 63108, USA			
PUBMED	Tel: 314 615 6979			
FEATURES	Fax: 314 615 5975			
source	Email: jbedell@oriongenomics.com			
	Plate: 529 row: k column: 22			
	Seq primer: T3 Reverse			
	Class: methylation filtered			
	High quality sequence stop: 590.			
	Location/Qualifiers			
	1. 590			
	/organism="Sorghum bicolor"			
	/mol_type="genomic DNA"			
	/cultivar="ATx623"			
	/db_xref="taxon:4558"			
	/clone="11135254"			
	/clone_1b="Sorghum methylation filtered library (LibID: 104)"			

/note="Organ: leaf; Vector: pBCSK(-); Site: 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 590;
Best Local Similarity 92.0%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
Db 353 AAAAAAAAAAGTCCCAATTCAGTTA 329

RESULT 5
CM091591/c 719 bp DNA linear GSS 28-OCT-2004
LOCUS 104 453 10998645 114 33028 085 Sorghum methylation filtered library
DEFINITION (LibID: 104) Sorghum bicolor genomic clone 10998645, genomic survey
sequence.

ACCESSION CM091591.1 GI:54768321
VERSION
KEYWORDS
SOURCE
ORGANISM

Sorghum bicolor (sorghum)
Sorghum bicolor
Bakayote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

1 (bases 1 to 719)
Bedell, J. A., Budiman, M. A., Nunberg, A., Citek, R. W., Robbins, D.,
Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K.,
McNamey, J., Smith, M., Holeman, H., Roe, B. A., Wiley, G., Korff, I. F.,
Rabinowitz, P. D., Lakey, N., McCombie, W. R., Jeddeloh, J. A. and
Martinsen, R. A.
Sorghum genome sequencing by methylation filtration
PLoS Biol. 3 (1), e13 (2005)
15660154

Contact: Bedell JA
Orion Genomics, LLC
4441 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975

Email: jbedell@oriongenomics.com
Plate: 453 row: k column: 21
Seq primer: M13/PUC Forward
Class: methylation filtered
High quality sequence stop: 719.
Location/Qualifiers

FEATURES

1..719
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/culivar="ATx623"
/db_xref="taxon:4558"
/clone_id="10998645"
/clone_lib="Sorghum methylation filtered library (LibID: 104)"
/note="Organ: leaf; Vector: pBCSK(-); Site: 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 719;
Best Local Similarity 92.0%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
Db 1 AAAAAAAAAAGTCCCAATTCAGATA 25

Db 132 AAAAAAAAAAGTCCCAATTCAGTTA 108

RESULT 6 777 bp mRNA linear EST 07-JAN-2005
CX451695
LOCUS UGI_XZG23543.rev NIH_XGC_tropae7 xenopus tropicalis cDNA clone
DEFINITION IMAGE:7539271 3', mRNA sequence.

ACCESSION CX451695
VERSION CX451695.1 GI:57267527
KEYWORDS
SOURCE
ORGANISM

REFERENCE

1 (bases 1 to 777)
Richardson, P., Lucas, S., Rokhsar, D., Deter, J. C., Ng, D. C.,
Brokstein, P. and Lindquist, E. A.
DOG Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: UGI_XZG23543.fwd
Contact: Lindquist, E. A., Richardson, P.
DOG Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LMD:
<http://image.llnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the UGI Clone Id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZG 0245 row: n column: 5
High quality sequence stop: 687
POLY=yes.
Location/Qualifiers

FEATURES

1..777
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7539271"
/tissue_type="whole embryo"
/dev_stage="Gastrula (st. 10.5-12.5)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
ElectroTen-Blue"
/clone_lib="NIH_XGC_tropae7"
/note="Vector: pCS108; Site: 1: SalI; Site 2: NotI.
Gastrula library constructed by Russell B. Fletcher in R.
Harland's lab using poly A RNA and oligo dt primers
(Invitrogen Superscript Plasmid System for cDNA synthesis
and cloning). SalI (5' end) -NotI (3' end) cDNA was
inserted into vector pCS108
(<http://mc.berkeley.edu/labs/harland/pages/plasmids.html>)
."

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 777;
Best Local Similarity 92.0%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
Db 704 AAAAAAAAAAGTCCCAATTCAGATA 728

RESULT 7
BI465187/c
LOCUS
DEFINITION BI465187 997 bp mRNA linear EST 21-AUG-2001
603208048P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5273722 5',
mRNA sequence.
ACCESSION BI465187
VERSION BI465187.1 GI:15255843
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 997)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: Miklos Palkevits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1AM1690 row: P column: 11
High quality sequence stop: 627.
Location/Qualifiers
1..997
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5273722"
/lab_host="DH10B"
/clone_1db="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI
(gfcgag); Oligo-gt primed using primer
5'-TTTTTTTTTTTTTTVA-3', size-selected for average
insert size 2.2 kb and normalized to ROT 5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN
Query Match 87.2%; Score 21.8; DB 3; Length 997;
Best Local Similarity 92.0%; Pred. No. 6.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
Db 811 AGAAGAAAGAGTCCCAATTCAGATA 787
|||||

RESULT 8
B08139
LOCUS
DEFINITION B08139 1000 bp DNA linear GSS 14-MAY-1997
F8J14-T7.1 IGF Arabidopsis thaliana genomic clone F8J14, genomic
survey sequence.
ACCESSION B08139
VERSION B08139.1 GI:20899419
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 1000)
AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other GSSs: F8J14-Sp6.1, F8J14-Sp6, F8J14-T7
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@genome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 564
High quality sequence stop: 694.
Location/Qualifiers
1..1000
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="F8J14"
/sex="thermaprodite"
/clone_1db="IGF"
/note="Vector: pBIOACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN
Query Match 87.2%; Score 21.8; DB 9; Length 1000;
Best Local Similarity 92.0%; Pred. No. 6.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
Db 313 AAAAAAAAAAGTCCCAATTCAGATA 337
|||||

RESULT 9
B12235
LOCUS
DEFINITION B12235 1003 bp DNA linear GSS 14-MAY-1997
F8J14-T7 IGF Arabidopsis thaliana genomic clone F8J14, genomic
survey sequence.
ACCESSION B12235
VERSION B12235.1 GI:2093356
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1003)
Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other GSSs: F8J14-Sp6.1, F8J14-T7.1, F8J14-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@genome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 578
High quality sequence stop: 669.
Location/Qualifiers
1..1003

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="F8J14"
/sex="hermaphrodite"
/note="Vector: BelobACII, Site_1: EcoRI, Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 1003;
Best Local Similarity 92.0%; Pred. No. 6.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 325 AAAAAAAAAAGTCCCAATTCAGAAA 349

RESULT 10
B12507 1030 bp DNA linear GSS 14-MAY-1997
LOCUS F19D14-Sp6.1 ICF Arabidopsis thaliana genomic clone F19D14, genomic
DEFINITION survey sequence.
ACCESSION B12507
VERSION B12507.1 GI:2093627
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.
1 (bases 1 to 1030)
Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
BAC End Sequences at ATGC
Unpublished (1997)
Other GSSes: F19D14-T7.1, F19D14-T7.2, F19D14-T7, F19D14-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends
High quality sequence start: 235
High quality sequence stop: 453.
Location/Qualifiers

FEATURES

1..1030
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="F19D14"
/sex="hermaphrodite"
/clone_1db="IGF"
/note="Vector: BelobACII, Site_1: EcoRI, Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN

Query Match 87.2%; Score 21.8; DB 9; Length 1030;
Best Local Similarity 92.0%; Pred. No. 6.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 411 AAAAAAAAAAGTCCCAATTCAGAAA 435

RESULT 11
Bj627421/c 331 bp mRNA linear EST 01-OCT-2003
LOCUS Bj627421 N1BB Mochii normalized Xenopus early gastrula library
DEFINITION Xenopus laevis cDNA clone XJ217p24.5', mRNA sequence.
ACCESSION B0627421
VERSION B0627421.1 GI:37273998
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 331)
Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-I, T. and
Kohara, Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadaeu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp
The information of this clone is available through the following
URL.
http://xenopus.nibb.ac.jp.
Location/Qualifiers

FEATURES

1..331
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8335"
/clone="XJ217p24"
/cissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_1db="N1BB Mochii normalized Xenopus early gastrula
library"

ORIGIN

Query Match 85.6%; Score 21.4; DB 3; Length 331;
Best Local Similarity 88.0%; Pred. No. 9.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 325 AAAAAAAAAAGTCCCAATTCAGATA 301

RESULT 12
BP704125/c 659 bp mRNA linear EST 19-JUL-2004
LOCUS BP704125 Oeada Taira anterior neuroectoderm (ANE) PCS105 cDNA
DEFINITION library Xenopus laevis cDNA clone XJ510m18ex 5', mRNA sequence.
ACCESSION BP704125
VERSION BP704125.1 GI:46052524
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 659)
Oeada, S., Kikayama, A., Ueno, N. and Taira, M.
Expression analysis of genes which are expressed in the anterior
neuroectoderm of Xenopus embryos
Unpublished (2004)
Contact: Masanori Taira
Department of Biological Sciences
Graduate School of Science, University of Tokyo; CREST, Japan
Science and Technology Corporation, Japan
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-03-5841-4434

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

Fax: 81-03-5841-4434
 Email: m_taira@biol.s.u-tokyo.ac.jp,
 URL: http://www.shigen.nig.ac.jp/brp/xenopus/est/
 Location/Qualifiers

FEATURES

source

1..659
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XLS10m18x"
 /tissue_type="anterior neuroectoderm"
 /dev_stage="late gastrula (stage 12.5)"
 /clone_lib="Osada Taira anterior neuroectoderm (AME)
 PCS105 cDNA library"

ORIGIN

Query Match 85.6%; Score 21.4; DB 3; Length 659;
 Best Local Similarity 88.0%; Pred. No. 9.7e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
 |||||
 Db 655 AAAAAAAAAAGTCCCAATTCAGATA 631

RESULT 13
 AGS57295/c 761 bp DNA linear GSS 23-DEC-2004
 LOCUS Mus musculus molossinus DNA, clone:MSMg01-475H15.TU, genomic survey
 DEFINITION sequence.
 ACCESSION AGS57295
 VERSION AGS57295.1 GI:48317993
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus (Japanese wild mouse)
 ORGANISM Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T.,
 Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaiki, K. and
 Shiroishi, T.
 Contribution of Asian mouse subspecies Mus musculus molossinus to
 genomic constitution of strain C57BL/6J, as defined by BAC-end
 sequence-SNP analysis
 Genome Res. 14 (12), 2439-2447 (2004)

JOURNAL PUBMED 15574823
 REFERENCE 2 (bases 1 to 761)
 AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 JOURNAL Direct Submission
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
 (E-mail: hattori@gsc.riken.jp, URL: http://bgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the mouse BAC library MSMg01. For BAC
 library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
 Tskuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1
 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp

COMMENT
 PRIMER
 Sequencing : TU
 LIBRARY Vector : PBACE3.6
 R Site 1 : EcoRI
 R Site 2 : EcoRI

FEATURES
 source Location/Qualifiers

1..761
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"

/clone="MSMg01-475H15.TU"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSMg01 Mouse Male BAC library"

ORIGIN

Query Match 85.6%; Score 21.4; DB 10; Length 761;
 Best Local Similarity 95.7%; Pred. No. 9.7e+02;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
 |||||
 Db 524 AAAAAAAAAAGTCCCAATTCAGA 502

RESULT 14
 CN640041/c 182 bp mRNA linear EST 30-SEP-2004
 LOCUS 261A11_554598 Douglas-fir cDNA library PmIFG_73-6 Pseudotsuga
 menziesii var. menziesii cDNA clone Df261-A11 5, mRNA sequence.
 DEFINITION
 ACCESSION CN640041
 VERSION CN640041.1 GI:47151118
 KEYWORDS EST.
 SOURCE Pseudotsuga menziesii var. menziesii
 ORGANISM Pseudotsuga menziesii var. menziesii
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pseudotsuga.

REFERENCE 1 (bases 1 to 182)
 Krutovsky, K.V., Trogiolo, M., Brown, G.R., Jeremstad, K.D. and
 Neale, D.B.
 Comparative mapping in the Pinaceae
 Genetics 168 (1), 447-461 (2004)

JOURNAL PUBMED 15454556
 COMMENT Contact: Krutovsky KV
 David Neale's lab
 Institute of Forest Genetics, USDA Forest Service, Pacific
 Southwest Research Station
 Department of Plant Sciences, University of California, One Shields
 Avenue, Davis, CA 95616, USA
 Tel: 1-530-752-8412
 Fax: 1-530-754-9366
 Email: krutovskyk@ucdavis.edu

PCR Primers
 FORWARD: M13 forward at 5' end
 Seg primer: M13 forward at 5' end
 High quality sequence stop: 182.

FEATURES

source Location/Qualifiers

1..182
 /organism="Pseudotsuga menziesii var. menziesii"
 /mol_type="mRNA"
 /varietal="menziesii"
 /db_xref="taxon:278161"
 /clone="Df261-A11"
 /tissue_type="entire seedlings"
 /dev_stage="one month old seedlings"
 /clone_lib="Douglas-fir cDNA library PmIFG_73-6"
 /note="Vector: pGEM-3Z; Site 1: EcoICRI; Site 2: EcoICRI;
 The cDNA library was obtained from one month old
 Douglas-fir seedlings. Total RNA was extracted from -5 g
 of ground tissue that consisted of the mixture of 10
 seedlings following the protocol described in Chang et
 al. (1993). mRNA was isolated from the total RNA using
 the PolyAtract mRNA Isolation System III (Promega 25300).
 Double stranded cDNA was prepared using the Universal
 Ribocloner cDNA Synthesis System (Promega C4360) and
 filtered through a Sephadryl S-400 column to remove low
 molecular weight products of less than 300 bp.
 Double-stranded blunt-end cDNA was ligated into the
 EcoICRI-cut dephosphorylated pGEM-3Z sequencing vector and
 electroporated into E. coli DH5 alpha. cDNA clones were
 sequenced by Integrated Genomics, Inc. (Chicago, USA)."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 182;
 Best Local Similarity 91.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
 158 AAAAAAAAAAGTACATTCAGAT 135

RESULT 15
 AL785515/c 214 bp mRNA linear EST 13-NOV-2003
 LOCUS AL785515 XGC-neurula Xenopus tropicalis cDNA clone TNeu070e21 5',
 DEFINITION mRNA sequence.
 AL785515
 ACCESSION AL785515.2 GI:38308469
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Xenopus tropicalis (western clawed frog)
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.

REFERENCE
 1 (bases 1 to 214)
 Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.
 Sanger Xenopus tropicalis EST project 2001 (11_2003)
 TITLE Unpublished (2003)
 JOURNAL On Jun 25, 2002 this sequence version replaced gi:21572129.
 COMMENT Contact: Taylor R
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn
 cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula.
 EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
 5' end and NotI at the 3' end.
 Vector: pCS107; Site 1: EcoRI; Site 2: NotI
 Host: Escherichia coli DH10B
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE ID: TNeu070e21.plc86
 Sequencing primer: SP6.
 Location/Qualifiers
 1..214
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:83564"
 /clone="TNeu070e21"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
 was oligo dt primed from 5ug of poly A+ RNA from neurula.
 EcoRI-NotI cut cDNA was then ligated into pCS107 with
 EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
 Query Match 83.2%; Score 20.8; DB 1; Length 214;
 Best Local Similarity 91.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
 172 AAAAAAAAAAGTCCATTCAGATA 149

RESULT 16
 BU615833/c 350 bp mRNA linear EST 01-OCT-2003
 LOCUS BU615833 NIBB Mochii normalized Xenopus early gastrula library
 DEFINITION Xenopus laevis cDNA clone XL174k21 5', mRNA sequence.
 BU615833
 ACCESSION BU615833.1 GI:37250862
 VERSION
 KEYWORDS
 EST.

SOURCE
 ORGANISM
 Xenopus laevis (African clawed frog)
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Xenopus.

REFERENCE
 1 (bases 1 to 350)
 Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-I, T. and
 Kohara, Y.
 TITLE Expressed genes in X. laevis embryo
 JOURNAL Unpublished (2001)
 COMMENT Contact: Tadao Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 The information of this clone is available through the following
 URL.
 http://xenopus.nibb.ac.jp.
 Location/Qualifiers
 1..350
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XL174k21"
 /tissue_type="whole embryo"
 /dev_stage="stage 10.5"
 /clone_lib="NIBB Mochii normalized Xenopus early gastrula
 library"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 3; Length 350;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
 343 AAAAAAAAAAGTCNCAATTCAGATA 319

RESULT 17
 BE101746 375 bp mRNA linear EST 13-JUN-2000
 LOCUS BE101746
 DEFINITION UR-R-B01-aps-h-04-0-UI.s1 UR-R-B01 Rattus norvegicus cDNA clone
 UR-R-B01-aps-h-04-0-UI 3', mRNA sequence.
 BE101746
 VERSION BE101746.1 GI:8493844
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus norvegicus (Norway rat)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.

REFERENCE
 1 (bases 1 to 375)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: Bento-soares@uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized hypothalamus library cDNA library Preparation: M.B.

Soares Lab Clone distribution: clones will be available through
Research Genetics (www.reagen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES

Location/Qualifiers

1..375
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-B01-aps-h-04-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-B01"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The library
(UI-R-B01) is a subcloned library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
reestc.eng.uiowa.edu. The subcloning has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_TISSUE=hypothalamus
TAG_LIB=UI-R-B01
TAG_SEQ=GATGC"

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 375;

Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAAAGTCCCAATTCAGAT 24

123 AAAAAAAAAAGTCCCAATTCAGAT 146

RESULT 18

LOCUS

CG329379 492 bp mRNA linear EST 04-JAN-2005
JGI_XZT68193.rev NIH_XGC_troptads Xenopus tropicalis cDNA clone

IMAGE:7787028 3', mRNA sequence.

CG329379

CG329379.1 GI:57065851

EST.

Xenopus tropicalis (western clawed frog)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodidae; Xenopus; Silurana.

1 (bases 1 to 492)

Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Xenopus tropicalis EST project

Unpublished (2004)

Other ESTs: JGI_XZT68193.fwd

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: chna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of

California, Berkeley: http://tropicalis.berkeley.edu/home

cDNA library Preparation: Richard M. Harland Laboratory, University

of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/LNL:

http://image.lnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone id and the direction of sequencing. The suffix '.rev'

indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0709 row: a column: 10
High quality sequence stop: 492
POLYA=yes.

FEATURES

source

Location/Qualifiers

1..492
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7787028"
/tissue_type="whole embryo"
/dev_stage="tadpole (ec. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
ElectroTen-Blue"
/clone_lib="NIH_XGC_troptads"
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
Superscript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector pCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 492;

Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 AAAAAAAAAAGTCCCAATTCAGATA 25

42 AAAAAAAAAAGTCCCAATTCAGATA 65

RESULT 19

LOCUS

CFI12690 514 bp mRNA linear EST 23-JUL-2003
Shultzoma05941 Rat lung airway and parenchyma cDNA libraries

Rattus norvegicus cDNA clone CA7244 5', mRNA sequence.

CFI12690

CFI12690.1 GI:33170842

EST.

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.

1 (bases 1 to 514)

Shultz, M.A., Zhang, L., Gu, Y.-Z., Baker, G.L., Pannuchi, M.V.,
Padua, A.M., Gurske, W.A., Morin, D., Penn, S.G., Jovanovich, S.B.,
Plopper, C.G. and Buckpitt, A.R.

Gene expression analysis in response to lung toxicants: I.

Sequencing and microarray development

Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)

Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)

Contact: Shultz MA

Dept. of Molecular Biosciences, School of Veterinary Medicine

University of California, Davis

1311 Harding Hall, One Shields Avenue, Davis, CA 95616, USA

Tel: 530 752 0793

Fax: 530 752 4698

Email: mshultz@ucdavis.edu

Average Phred score is 20 or better. All poor quality data (Phred <

20) and vector/linker sequence has been removed.

High quality sequence stop: 514.

Location/Qualifiers

FEATURES

source

1. 514
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="CA7244"
/sex="male"
/tissue_type="airway or parenchyma"
/dev_stage="adult"
/clone_lib="Rat lung airway and parenchyma cDNA libraries"
/note="Organ: lung; Vector: pGEM-11zf(-); Site 1: Eco RI; Site 2: Not I; mRNA was isolated from microdissected rat lung airways and parenchyma tissues."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 514;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
|||||
513 AAAAAAAAAAGTCCCAATTCAGAT 490

Db

RESULT 20
CX472437 514 bp mRNA linear EST 10-JAN-2005
LOCUS JGI_XZG5560.rev NIH XGC tropoGas7 Xenopus tropicalis cDNA clone
DEFINITION IMAGE:7523207 3', mRNA sequence.
CX472437
VERSION CX472437.1 GI:57363365
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 514)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J. C., Ng, D. C., Brokstein, P. and Lindquist, E. A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI_XZG5560.fed
Contact: Lindquist, E. A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdaa@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/BLNI:
<http://image.lnl.gov>
Naming Conventions: EST name is generated by the concatenation of the JGI Clone id and the direction of sequencing. The suffix 'rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
poly-A: Based upon the presence of a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZG 0057 row: P column: 21
High quality sequence stop: 380
POLYA=yes

FEATURES

Location/Qualifiers

1. 514
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7523207"
/tissue_type="whole embryo"

source

/dev_stage="Gastrula (st. 10.5-12.5)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electrogen-Blue"
/clone_lib="NIH XGC tropoGas7"
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI; Gastrula library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS108
(<http://ncb.berkeley.edu/labs/harland/pages/plasmids.html>)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 514;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
42 AAAAAAAAAAGTCCCTTCAGATA 65

Db

RESULT 21
AL969417/c 523 bp mRNA linear EST 05-DEC-2003
LOCUS AL969417 XGC-gastrula Xenopus tropicalis cDNA clone TGae092h02 5',
DEFINITION mRNA sequence.
AL969417
VERSION AL969417.2 GI:39027481
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 523)
Croning, M. D. R., Ashurst, J. L., Taylor, R., Zorn, A. M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
On Nov 27, 2002 this sequence version replaced gi:25793012.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: TGae092h02.p1kSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-Blue.
Location/Qualifiers

1. 523
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGae092h02"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-Blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 83.2%; Score 20.8; DB 1; Length 523;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;

Matches 22: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
 |||||
 484 AAAAAAAAAAGTCCCAATTCAGATA 461

RESULT 22
 BP709610/c 529 bp mRNA linear EST 25-OCT-2004
 LOCUS BP709610 Osada Taira anterior neuroectoderm (ANE) PCS105 cDNA
 DEFINITION library Xenopus laevis cDNA clone XL498h18ex 5', mRNA sequence.
 BP709610
 ACCESSION BP709610.2 GI:54630123
 VERSION EST.
 KEYWORDS Xenopus laevis (African clawed frog)
 SOURCE Xenopus laevis
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Xenopus.
 1 (bases 1 to 529)
 Osada,S., Kitayama,A., Ueno,N. and Taira,M.
 Expression analysis of genes which are expressed in the anterior
 neuroectoderm of Xenopus embryos
 Unpublished (2004)
 JOURNAL On Jul 19, 2004 this sequence version replaced gi:46058009.
 COMMENT Contact: Masanori Taira
 Department of Biological Sciences
 Graduate School of Science, University of Tokyo; CREST, Japan
 Science and Technology Corporation, Japan
 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
 Tel: 81-03-5841-4434
 Fax: 81-03-5841-4434
 Email: m.taira@biol.s.u-tokyo.ac.jp,
 URL: http://www.shigen.nig.ac.jp/hbrrp/xenopus/est/.

FEATURES
 source
 1..529
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XL498h18ex"
 /tissue_type="anterior neuroectoderm"
 /dev_stage="late gastrula (stage 12.5)"
 /clone_lib="osada Taira anterior neuroectoderm (ANE)
 PCS105 cDNA library"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 3; Length 529;
 Best Local Similarity 88.0%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
 |||||
 522 AAAAAAAAAAGTCCCAATTCAGATA 498

RESULT 23
 CX474359 551 bp mRNA linear EST 10-JAN-2005
 LOCUS JGI.XZG49319.rev NIH XGC tropogastr Xenopus tropicalis cDNA clone
 DEFINITION IMAGE:7564243 3', mRNA sequence.
 CX474359
 ACCESSION CX474359.1 GI:57367185
 VERSION EST.
 KEYWORDS Xenopus tropicalis (western clawed frog)
 SOURCE Xenopus tropicalis
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 551)
 Richardson,P., Lucas,S., Rokhsar,D., Dettter,J.C., Ng,D.C.,
 Brokstein,P. and Lindquist,E.A.
 DOE Joint Genome Institute Xenopus tropicalis EST project

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

JOURNAL Unpublished (2004)
 COMMENT Other ESTs: JGI.XZG49319.fwd
 Contact: Lindquist,E.A., Richardson,P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 296 5600
 Fax: 925 296 5710
 Email: cdna@jgi-psf.org
 Tissue Procurement: Richard M. Harland laboratory, University of
 California, Berkeley: http://tropicalis.berkeley.edu/home
 cDNA Library Preparation: Richard M. Harland laboratory, University
 of California, Berkeley
 DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
 Clone Distribution: I.M.A.G.E. Consortium/LNL:
 http://image.lnl.gov
 Naming Conventions: EST name is generated by the concatenation of
 the JGI Clone id and the direction of sequencing. The suffix 'rev'
 indicates a reverse sequencing read of the insert. It does not
 necessarily reflect the orientation of the insert.
 Poly-A: Based upon the presence of a run of 14 or more T residues
 at the beginning of the sequence, this clone was polyadenylated.
 The resulting Poly-T sequence has been removed.
 Plate: XZG 0513 row: n column: 17
 High quality sequence stop: 551
 POLYA=yes

FEATURES
 source
 1..551
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="IMAGE:7564243"
 /tissue_type="whole embryo"
 /dev_stage="gastrula (st. 10.5-12.5)"
 /lab_host="E. coli XL1-Blue derivative, Stratagene
 ElectroTen-Blue"
 /clone_lib="NIH XGC tropogastr"
 /note="Vector: PCS108; Site 1: SalI; Site 2: NotI;
 Gastrula library constructed by Russell B. Fletcher in R.
 Harland's lab using poly A RNA and oligo dt primers
 (Invitrogen Superscript Plasmid System for cDNA Synthesis
 and Cloning). SalI (5' end) -NotI (3' end) cDNA was
 inserted into vector PCS108
 (http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 8; Length 551;
 Best Local Similarity 91.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
 |||||
 42 AAAAAAAAAAGTCCCAATTCAGATA 65

RESULT 24
 CN949747 560 bp mRNA linear EST 08-JUN-2004
 LOCUS XXXPI55 Nicotiana tabacum Pollen PCR-based subtractive library
 DEFINITION Nicotiana tabacum cDNA clone Peg134BS, mRNA sequence.
 CN949747
 ACCESSION CN949747.1 GI:48431337
 VERSION EST.
 KEYWORDS Nicotiana tabacum (common tobacco)
 SOURCE Nicotiana tabacum
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamiales; Solanales; Solanaceae; Nicotiana.
 1 (bases 1 to 560)
 Shary,S., Kumar,R. and Guha-Mukherjee,S.
 Isolation and expression studies of early genes involved in pollen
 development
 Unpublished (2004)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT Contact: Semrajit Shary
Plant Molecular Biology
Jawaharlal Nehru University
School of Life Sciences, New Delhi -110067 India
Tel: 26187175
Email: shelli8977@yahoo.com
Insert Length: 560 Std Error: 0.00.
Location/Qualifiers

FEATURES
Source
1. .560
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Xanthi.1"
/db_xref="taxon:4097"
/clone="Peg134BS"
/tissue_type="Pollen"
/lab_host="E. coli"
/clone_lib="Nicotiana tabacum Pollen PCR-based subtractive library"
/note="Vector: p-GEMT Easy"

ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 560;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
|||||
Db 186 AAAAAAAAAATCTCAATTCAGAT 209
|||||

RESULT 25
CX418273 567 bp mRNA linear EST 06-JAN-2005
DEFINITION JGI XZG65446.rev NIH XGC tropGae7 Xenopus tropicalis cDNA clone
IMAGE:7579555 3', mRNA sequence.
CX418273
ACCESSION CX418273.1 GI:57198976
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoes; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 567)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI XZG65446.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNLN:
<http://image.lnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZG 0681 row: 1 column: 17
High quality sequence stop: 554
POLYA=Yes.
Location/Qualifiers

source
1. .567
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7579555"
/tissue_type="whole embryo"
/dev_stage="Gastrula (st. 10.5-12.5)"
/lab_host="E. coli XL1-Blue derivative, StrataGene
ElectroTen-Blue"
/note="NIH XGC tropGae7"
/note="Vector: pCS108, Site 1: SalI, Site 2: NotI;
Gastrula library constructed by Russell B. Fletcher in R.
Harland's lab using poly A RNA and oligo dt primers
(Invitrogen Superscript Plasmid System for cDNA Synthesis
and Cloning). SalI (5' end) -NotI (3' end) cDNA was
inserted into vector pCS108
(<http://mcb.berkeley.edu/labs/harland/pages/plasmids.html>)
."

ORIGIN
Query Match 83.2%; Score 20.8; DB 8; Length 567;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
|||||
Db 42 AAAAAAAAAAGTCCCAATTCAGATA 65
|||||

RESULT 26
CX450676 567 bp mRNA linear EST 07-JAN-2005
DEFINITION JGI XZG22958.rev NIH XGC tropGae7 Xenopus tropicalis cDNA clone
IMAGE:7538454 3', mRNA sequence.
CX450676
ACCESSION CX450676.1 GI:57266508
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoes; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 567)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI XZG22958.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNLN:
<http://image.lnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZG 0237 row: 1 column: 4
High quality sequence stop: 519
POLYA=Yes.
Location/Qualifiers
1. .567

FEATURES

FEATURES
source

/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7538454"
/tissue_type="whole embryo"
/dev_stage="Gastrula (st. 10.5-12.5)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
ElectroTen-Blue"
/clone_id="NH_XGC_tropGas7"
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI;
Gastrula library constructed by Russell B. Fletcher in R.
Harland's lab using poly A RNA and oligo dt primers
(Invitrogen Superscript Plasmid System for cDNA Synthesis
and Cloning). SalI (5' end) -NotI (3' end) cDNA was
inserted into vector pCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 567;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAAGTCCATTGAGATA 25
Db 42 AAAAAAAAAAGTCCATTGAGATA 65

RESULT 27

CK417651 578 bp mRNA linear EST 06-JAN-2005
LOCUS UGI_XZG65088.rev NIH_XGC_tropGas7 Xenopus tropicalis cDNA clone
DEFINITION IMAGE:7579273 3, mRNA sequence.
ACCESSION CK417651
VERSION CK417651.1 GI:57198354
KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 578)

REFERENCE Richardson, P., Lucas, S., Rokhsar, D., Dettler, J. C., Ng, D. C.,
Brokstein, P. and Lindquist, E. A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNAL Unpublished (2004)

COMMENT Other ESTs: JGI XZG65088.fwd
Contact: Lindquist, E. A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710

Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LNL:
http://image.lnl.gov

Naming Conventions: EST name is generated by the concatenation of
the JGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZG 0677 row: p column: 23
High quality sequence stop: 529
POLYA=yes.

FEATURES

Location/Qualifiers
1..578
/organism="Xenopus tropicalis"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 578;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAAGTCCATTGAGATA 25
Db 42 AAAAAAAAAAGTCCATTGAGATA 65

RESULT 28

CR429411 589 bp mRNA linear EST 17-JUN-2004
LOCUS CR429411
DEFINITION CR429411 XGC-tailbud Xenopus tropicalis cDNA clone TTba068019 3',
mRNA sequence.
ACCESSION CR429411
VERSION CR429411.1 GI:48922820
KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 589)

REFERENCE Croning, M. D. R., Ashurst, J. L., Taylor, R., Garrett, N. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)

COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE ID: TTba068019.q1kT7

This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: T7.
Location/Qualifiers
1..589
/organism="Xenopus tropicalis"

FEATURES

Location/Qualifiers
1..589
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTba068019"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_id="XGC-tailbud"
/note="Vector: pCS107; Site 1: EcoRI, site 2: NotI; cDNA
was oligo dt primed from 5' end of poly A+ RNA from tailbud.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 589;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAAGTCCATTGAGATA 25

Db 82 AAAAAAAAAAGTCCATTACAGATA 105

RESULT 29
LOCUS CX496283
DEFINITION JGI_XZG38090.rev NIH_XGC_tropGad5 Xenopus tropicalis cDNA clone
IMAGE:7553229 3', mRNA sequence.
ACCESSION CX496283
VERSION CX496283.1 GI:57412758
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
REFERENCE
AUTHORS Richardson,P., Lucas,S., Rokhsar,D., Dettler,J.C., Ng,D.C.,
Brokstein,P. and Lindquist,E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNAL Unpublished (2004)
COMMENT Other ESTs: JGI_XZG38090.fwd
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNLN:
<http://image.lnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the UGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZG 0397 row: c column: 19
High quality sequence stop: 483
POLYA=yes.

FEATURES

Source

Location/Qualifiers
1..614
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7553229"
/issue_type="whole embryo"
/dev_stage="Gastrula (st. 10.5-12.5)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_lib="NIH_XGC_tropGad5"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI;
Gastrula library constructed by Russell B. Fletcher in R.
Harland's lab using poly A RNA and oligo dt primers (Invitrogen
Superscript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was
inserted into vector PCS108
(<http://mc.berkeley.edu/labs/harland/pages/plasmids.html>)
."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 614;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGTCCATTACAGATA 25
|||||

Db 417 AAAAAAAAAAGTCCATTACAGATA 440

RESULT 30
LOCUS CX413722
DEFINITION JGI_XZT4941.rev NIH_XGC_tropGad5 Xenopus tropicalis cDNA clone
IMAGE:75683342 3', mRNA sequence.
ACCESSION CX413722
VERSION CX413722.1 GI:57194425
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
REFERENCE
AUTHORS Richardson,P., Lucas,S., Rokhsar,D., Dettler,J.C., Ng,D.C.,
Brokstein,P. and Lindquist,E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNAL Unpublished (2004)
COMMENT Other ESTs: JGI_XZT4941.fwd
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LNLN:
<http://image.lnl.gov>
Naming Conventions: EST name is generated by the concatenation of
the UGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZT 0049 row: j column: 12
High quality sequence stop: 552
POLYA=yes.

FEATURES

Source

Location/Qualifiers
1..655
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7583342"
/issue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_lib="NIH_XGC_tropGad5"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
Superscript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector PCS108
(<http://mc.berkeley.edu/labs/harland/pages/plasmids.html>)
."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 655;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGTCCATTACAGATA 25
|||||
Db 43 AAAAAAAAAAGTCCATTACAGATA 66
|||||

RESULT 31
DR902299
LOCUS
DEFINITION JG1 XZT61174.rev NIH XGC tropiclaids Xenopus tropicalis cDNA clone
IMAGE:7635991 3', mRNA sequence.
ACCESSION DR902299
VERSION DR902299.1 GI:71591551
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 655)
AUTHORS Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,
Broksrein, P. and Lindquist, E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNAL Unpublished (2004)
COMMENT Other ESTs: JG1_XZT61174.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: ghaadgi@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LLNL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the JGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: XZT 0637 row: 1 column: 5
High quality sequence stop: 554
POLYAs:Yes

FEATURES
source
1..655
Location/Qualifiers
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7635991"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_lib="NIH_XGC_tropTad5"
/note="Vector: PCS108; Site_1: SalI; Site_2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
Superscript Plasmid System for cDNA Synthesis and
Cloning) SalI (5' end) -NotI (3' end) cDNA was inserted
into vector PCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

ORIGIN
Query Match 83.2%; Score 20.8; DB 8; Length 655;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 42 AAAAAAAAAAGTCCCAATTCAGATA 65

RESULT 32
BU418890/c
LOCUS
DEFINITION 603959342F1 CSRBORBN09 Gallus gallus cDNA clone CHES1932c11 5', mRNA
sequence.
ACCESSION BU418890
VERSION BU418890.1 GI:25911561
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 664)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken CDNA
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1..664
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="layer and broiler"
/db_xref="taxon:9031"
/clone="CHES1932c11"
/sex="Male and female"
/tissue_type="Chondrocytes isolated from growth plate
cartilage"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSRBORBN09"
/note="Vector: pBluescript II KS(+); Site_1: EcoRI;
Site_2: NotI; This normalized library was constructed from
1 million independent clones. cDNA synthesis was initiated
using an oligo(dT) primer, using methylated C in the first
strand synthesis reaction. Following this first strand
reaction, double-stranded cDNA was blunt-ended, ligated to
NotI adapters, digested with EcoRI, size-selected, and
cloned into the NotI and EcoRI compatible sites of a
custom modified MCS of the pBluescript (KS+) vector. The
library was normalized in 2 rounds using conditions
adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
Bonaldo et al., Genome Research 6 (1996): 791, except that
a significantly longer reannealing hybridization was
used."

ORIGIN
Query Match 83.2%; Score 20.8; DB 5; Length 664;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 487 AAAAAAAAAAGTCCCAATTCAGAT 464

RESULT 33
CR560919/c
LOCUS
DEFINITION 698 bp mRNA linear EST 19-JUL-2004
CR560919 XGC-tailbud-head Xenopus tropicalis cDNA clone ThdA002c02
3', mRNA sequence.
ACCESSION CR560919
VERSION CR560919.1 GI:50390996

KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

REFERENCE
AUTHORS Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: THDA002c02.g1kT7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: T7.
Location/Qualifiers
1..698
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA002c02"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_1lb="XGC-tailbud-head"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from tailbud
head. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 698;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
DB 277 AAAAAAAAAAGTTCATTCAGAT 254

RESULT 34
LOCUS CR560918 710 bp mRNA linear EST 19-JUN-2004
DEFINITION CR560918 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA002c02
5', mRNA sequence.
ACCESSION CR560918
VERSION CR560918.1 GI:50390995
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

REFERENCE
AUTHORS Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: THDA002c02.plkSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: SP6.
Location/Qualifiers
1..710
/organism="Xenopus tropicalis"
/mol_type="mRNA"

/db_xref="taxon:8364"
/clone="THDA002c02"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_1lb="XGC-tailbud-head"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from tailbud
head. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 710;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
DB 461 AAAAAAAAAAGTTCATTCAGAT 484

RESULT 35
LOCUS CR413014 711 bp mRNA linear EST 13-JUN-2004
DEFINITION CR413014 XGC-tailbud Xenopus tropicalis cDNA clone TTBA063b04 5',
mRNA sequence.
ACCESSION CR413014
VERSION CR413014.1 GI:48681261
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

REFERENCE
AUTHORS Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TTBA063b04.plkSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: SP6.
Location/Qualifiers
1..711
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTBA063b04"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_1lb="XGC-tailbud"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from tailbud.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 711;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
DB 572 AAAAAAAAAAGTTCATTCAGAT 595

RESULT 36
LOCUS CX401299 715 bp mRNA linear EST 06-JAN-2005

DEFINITION JGI XZT49120 rev NIH XGC_troptrads Xenopus tropicalis cDNA clone
IMAGE:7624194 3', mRNA sequence.

ACCESSION CX401299
VERSION CX401299.1 GI:57181989

KEYWORDS EST.
Xenopus tropicalis (western clawed frog)

SOURCE
ORGANISM Xenopus tropicalis

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 715)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C., Brokerstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI XZT49120.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/LLNL: <http://image.llnl.gov>
Naming Conventions: EST name is generated by the concatenation of the JGI clone id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Small Insert: Based upon one or more sequencing reads of this clone where vector sequence was present at both ends, this clone has been determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0509 .row: P column: 16
High quality sequence stop: 637
POLYA=Yes.

FEATURES
source .
Location/Qualifiers
1..715
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7624194"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electroten-Blue"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector PCS108
(<http://mcb.berkeley.edu/labs/harland/pages/plasmids.html>)

ORIGIN
Query Match 83.2%; Score 20.8; DB 8; Length 715;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCATTTCAGATA 25
|||||
Db 443 AAAAAAAAAAGTCCATTTCAGATA 466

RESULT 37

CR565819/c
LOCUS 721 bp mRNA linear EST 19-JUL-2004
DEFINITION CR565819 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA013e08 3', mRNA sequence.

ACCESSION CR565819
VERSION CR565819
KEYWORDS EST.
Xenopus tropicalis (western clawed frog)

SOURCE
ORGANISM Xenopus tropicalis

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 721)
Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001
Unpublished (2004)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: THDA013e08.g1kT7
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.
Seq primer: T7.

FEATURES
source
Location/Qualifiers
1..721
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA013e08"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Escherichia coli DH10b."
/clone_lib="XGC-tailbud-head"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from tailbud head. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 721;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTTCAGAT 24
|||||
Db 276 AAAAAAAAAAGTCCATTTCAGAT 253

RESULT 38
LOCUS 734 bp mRNA linear EST 07-FEB-2005
DEFINITION JGI CAANS826 rev NIH XGC_troptr4d Xenopus tropicalis cDNA clone
IMAGE:7691467 3', mRNA sequence.

ACCESSION CX916844
VERSION CX916844
KEYWORDS EST.
Xenopus tropicalis (western clawed frog)

SOURCE
ORGANISM Xenopus tropicalis

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 734)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C., Brokerstein, P. and Lindquist, E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI CAANS826.fwd
Contact: Lindquist, E.A., Richardson, P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710

Email: cdna@jgi-psf.org
 Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,
 University of California, Berkeley:
<http://tropicalis.berkeley.edu/home>)
 cDNA Library Preparation: DOE Joint Genome Institute:
<http://www.jgi.doe.gov>
 DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
 Clone Distribution: I.M.A.G.E. Consortium/LNL:
<http://image.lnl.gov>
 Naming Conventions: EST name is generated by the concatenation of
 the JGI Clone id and the direction of sequencing. The suffix '.rev'
 indicates a reverse sequencing read of the insert. It does not
 necessarily reflect the orientation of the insert.
 Poly-A: Based upon the presence of a run of 14 or more T residues
 at the beginning of the sequence, this clone was polyadenylated.
 The resulting Poly-T sequence has been removed.
 Plate: CAAN 0061 row: C column: 17
 High quality sequence stop: 634
 POLY-A: yes.

FEATURES

source

1..734
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="IMAGE:7691467"
 /sex="male"
 /tissue_type="Testes"
 /dev_stage="Adult"
 /lab_host="Electromax DH10B"
 /clone_1lib="NIH_XGC_tropTet4"
 /note="Vector: PCMVSPORT6; Site_1: SalI, Site_2: NotI;
 This library was made from dt primed cDNA and cloned into
 Invitrogen PCMVSPORT6 vector. The work was done at DOE
 Joint Genome Institute. Poly A RNA were primed with 5'
 GACTAGTTCTAGATCGGAG CGGCGCCCTTTT TTTT TTTT 3'. cDNA
 were ligated to SalI adapter (5' TCACACCGACGCTCCG and
 5' CGAGCGCTGGG), digested with NotI, size fractionated in
 1.1% agarose gel electrophoresis and ligated into NotI and
 SalI digested PCMVSPORT6 vector."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 734;
 Best Local Similarity 91.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCATTACAGATA 25
 443 AAAAAAAAAAGTCCATTACAGATA 466

RESULT 39
 BX727946 738 bp mRNA linear EST 18-NOV-2003
 LOCUS BX727946 XGC-tadpole Xenopus tropicalis cDNA clone TTPA071h19 3',
 DEFINITION mRNA sequence.
 ACCESSION BX727946
 VERSION BX727946.1 GI:38400687
 KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 738)
 Reference: Cronling M.D.R., Ashurst J.L., Taylor R., Zorn A.M. and Rogers J.
 Sanger Xenopus tropicalis EST project 2001 (1_2003)
 JOURNAL Unpublished (2003)
 COMMENT Contact: Cronling MDR
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE_ID: TTPA071h19.g1kat7

Sequencing primer: T7
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Nigel Garrett.
 cDNA was oligo dt primed from 5ug of poly A+ RNA from tadpole
 embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
 EcoRI at the 5' end and NotI at the 3' end.
 Vector: pCS107; Site_1: EcoRI; Site_2: NotI
 Host: Escherichia coli DH10B.
 Location/Qualifiers

FEATURES

source

1..738
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TTPA071h19"
 /dev_stage="tadpole (stage 35-40)"
 /lab_host="E. coli DH10B"
 /clone_1lib="XGC-tadpole"
 /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
 was oligo dt primed from 5ug of poly A+ RNA from tadpole
 embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
 with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 738;
 Best Local Similarity 91.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCATTACAGATA 25
 488 AAAAAAAAAAGTCCATTACAGATA 511

RESULT 40
 CX359920 757 bp mRNA linear EST 05-JAN-2005
 LOCUS CX359920
 DEFINITION JGI_X2T958.fwd NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone
 IMAGE:7581097 5', mRNA sequence.
 ACCESSION CX359920
 VERSION CX359920.1 GI:57128479
 KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 757)
 Reference: Richardson P., Lucas S., Rokhsar D., Dettner J.C., Ng D.C.,
 Brokstein P. and Lindquist E.A.
 DOE Joint Genome Institute Xenopus tropicalis EST project
 Unpublished (2004)
 JOURNAL Other ESTs: JGI_X2T958.rev
 Contact: Lindquist E.A., Richardson P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 296 5600
 Fax: 925 296 5710
 Email: cdna@jgi-psf.org
 Tissue Procurement: Richard M. Harland Laboratory, University of
 California, Berkeley: <http://tropicalis.berkeley.edu/home>
 cDNA Library Preparation: Richard M. Harland Laboratory, University
 of California, Berkeley
 DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
 Clone Distribution: I.M.A.G.E. Consortium/LNL:
<http://image.lnl.gov>
 Naming Conventions: EST name is generated by the concatenation of
 the JGI Clone id and the direction of sequencing. The suffix '.fwd'
 indicates a forward sequencing read of the insert. It does not
 necessarily reflect the orientation of the insert.
 Plate: XZT 0009 row: 1 column: 23
 High quality sequence stop: 730.
 Location/Qualifiers
 1..757
 /organism="Xenopus tropicalis"

```

/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7598032"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_id="NH_XGC_tropTad5"
/notice="Vector: pCS108; Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector pCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)
"

```

ORIGIN

```

Query Match      83.2%; Score 20.8; DB 8; Length 757;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
    |||||
Db 391 AAAAAAAAAAGTTCATTCAGAT 414

```

```

RESULT 41
CX387336/c 778 bp mRNA linear EST 05-JAN-2005
DEFINITION JGI_XZT21463.rev NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone
ACCESSION CX387336
VERSION IMAGE:7598032.3, mRNA sequence.
KEYWORDS CX387336.1 GI:57155893
SOURCE EST.

```

```

ORGANISM Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.

```

```

REFERENCE 1 (bases 1 to 778)
AUTHORS Richardson,P., Lucas,S., Rokhsar,D., Dettler,J.C., Ng,D.C.,
Birkenstein,P. and Lindquist,E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNML Unpublished (2004)
COMMENT Other ESTs: JGI_XZT21463.fwd
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LLNL:
http://image.llnl.gov

```

Naming Conventions: EST name is generated by the concatenation of the JGI Clone id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.

Poly-A: Based upon the presence of a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated. The resulting Poly-T sequence has been removed.

Plate: XZT 0221 row: n column: 14

High quality sequence stop: 693

POLYA=Yes.

FEATURES

```

Source location/Qualifiers
1..778
/organism="Xenopus tropicalis"
/mol_type="mRNA"

```

```

/db_xref="taxon:8364"
/clone="IMAGE:7598032"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"
/clone_id="NH_XGC_tropTad5"
/notice="Vector: pCS108; Site 1: SalI; Site 2: NotI; Tadpole
library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector pCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)
"

```

ORIGIN

```

Query Match      83.2%; Score 20.8; DB 8; Length 778;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
    |||||
Db 250 AAAAAAAAAAGTTCATTCAGAT 227

```

```

RESULT 42
CX391469 793 bp mRNA linear EST 06-JAN-2005
DEFINITION JGI_XZT38998.fwd NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone
ACCESSION CX391469
VERSION IMAGE:7615232.5, mRNA sequence.
KEYWORDS CX391469.1 GI:57172149
SOURCE EST.

```

```

ORGANISM Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.

```

```

REFERENCE 1 (bases 1 to 793)
AUTHORS Richardson,P., Lucas,S., Rokhsar,D., Dettler,J.C., Ng,D.C.,
Birkenstein,P. and Lindquist,E.A.
TITLE DOE Joint Genome Institute Xenopus tropicalis EST project
JOURNML Unpublished (2004)
COMMENT Other ESTs: JGI_XZT38998.rev
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LLNL:
http://image.llnl.gov

```

Naming Conventions: EST name is generated by the concatenation of the JGI Clone id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.

Plate: XZT 0405 row: k column: 6

High quality sequence stop: 790.

FEATURES

```

Source location/Qualifiers
1..793
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7615232"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene

```

ElectroTen-Blue"
/clone_id="NIH_XGC_tropTad5"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector PCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 793;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTCCATTACAGAT 24
Db 679 AAAAAAAAAAGTCCATTACAGAT 702

RESULT 43

LOCUS BX738214 795 bp mRNA linear EST 18-NOV-2003
DEFINITION BX738214 XGC-tadpole Xenopus tropicalis cDNA clone TTPA071h22.3',
mRNA sequence.

ACCESSION BX738214 GI:38410955
VERSION BX738214
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 795)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK

Email: tropesanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TTPA071h22.q1kat7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.

cDNA was oligo dt primed from 5ug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into PCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: PCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli DH10B.

Location/Qualifiers

1..795
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTPA071h22"
/dev_stage="tadpole (stage 35-40)"
/lab_host="E. coli DH10B"
/clone_lib="XGC-tadpole"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into PCS107
with EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 795;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 AAAAAAAAAAGTCCATTACAGATA 25
|||||

Db 480 AAAAAAAAAAGTCCATTACAGATA 503

RESULT 44

LOCUS CX329053/c 802 bp mRNA linear EST 04-JAN-2005
DEFINITION JGI_XZT67960.rev NIH XGC_tropTad5 xenopus tropicalis cDNA clone
IMAGE:7787016.3', mRNA sequence.

ACCESSION CX329053
VERSION CX329053
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 802)
Richardson,P., Lucas,S., Rokhsar,D., Dettner,J.C., Ng,D.C.,
Brokstein,P. and Lindquist,E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI_XZT67960.fwd
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home
cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LMN:
http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of
the JGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.

Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0705 row: P column: 22
High quality sequence stop: 716
POLYA=yes.

Location/Qualifiers

1..802
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7787016"
/tissue_type="whole embryo"
/dev_stage="tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
ElectroTen-Blue"
/clone_lib="NIH_XGC_tropTad5"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's
lab using poly A RNA and oligo dt primers (Invitrogen
SuperScript Plasmid System for cDNA Synthesis and
Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
into vector PCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 802;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTGAGAT 24
 |||||
 Db 250 AAAAAAAAAAGTTACAATTGAGAT 227

RESULT 45
 CX362328
 LOCUS CX362328
 DEFINITION UGI_XZT42336.fwd NIH_XGC_tropIads xenopus tropicalis cDNA clone
 IMAGE:7618033 5', mRNA sequence.
 ACCESSION CX362328
 VERSION CX362328.1 GI:57130887
 KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 813)
 Richardson, P., Lucas, S., Rokhsar, D., Detter, J.C., Ng, D.C.,
 Brockstein, P. and Lindquist, E.A.
 DOE Joint Genome Institute Xenopus tropicalis EST project
 Unpublished (2004)
 Other ESTs: UGI_XZT42336.rev
 Contact: Lindquist, E.A., Richardson, P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 296 5600
 Fax: 925 296 5710
 Email: cgha@jgi-psf.org
 Tissue Procurement: Richard M. Harland Laboratory, University of
 California, Berkeley: http://tropicalis.berkeley.edu/home
 cDNA Library Preparation: Richard M. Harland Laboratory, University
 of California, Berkeley
 DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
 Clone Distribution: I.M.A.G.E. Consortium/LLNL:
 http://image.llnl.gov
 Naming Conventions: EST name is generated by the concatenation of
 the UGI clone id and the direction of sequencing. The suffix '.fwd'
 indicates a forward sequencing read of the insert. It does not
 necessarily reflect the orientation of the insert.
 Plate: XZT 0441 row: 0 column: 23
 High quality sequence stop: 771.
 Location/Qualifiers
 1..813
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="IMAGE:7618033"
 /tissue_type="whole embryo"
 /dev_stage="Tadpole (st. 36-41)"
 /lab_host="E. coli XL1-Blue derivative, Stratagene
 Electrolten-Blue"
 /clone_lib="NIH_XGC_tropIads"
 /note="Vector: pCS108; Site 1: SalI; Site 2: NotI; Tadpole
 library constructed by Russell B. Fletcher in R. Harland's
 lab using poly A RNA and oligo dt primers (Invitrogen
 SuperScript Plasmid System for cDNA Synthesis and
 Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted
 into vector pCS108
 (http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 8; Length 813;
 Best Local Similarity 91.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTGAGAT 24
 |||||
 Db 206 AAAAAAAAAAGTTACAATTGAGAT 229

RESULT 46
 CR569461
 LOCUS CR569461
 DEFINITION CR569461 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA030103
 5', mRNA sequence.
 ACCESSION CR569461
 VERSION CR569461.1 GI:50456887
 KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 815)
 Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.
 Sanger Xenopus tropicalis EST project 2001 (2004)
 Unpublished (2004)
 Contact: Croning MDR
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE_ID: THDA030103.plkasp6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Nigel Garrett.
 Seq primer: SP6.
 Location/Qualifiers
 1..815
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="THDA030103"
 /dev_stage="tailbud head (stage 28-30)"
 /lab_host="Escherichia coli DH10B."
 /clone_lib="XGC-tailbud-head"
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
 was oligo dt primed from 5'ug of poly A+ RNA from tailbud
 head. EcoRI-NotI cut cDNA was then ligated into pCS107
 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
 Query Match 83.2%; Score 20.8; DB 7; Length 815;
 Best Local Similarity 91.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTGAGAT 24
 |||||
 Db 765 AAAAAAAAAAGTTACAATTGAGAT 788

RESULT 47
 CR413015/c
 LOCUS CR413015
 DEFINITION CR413015 XGC-tailbud Xenopus tropicalis cDNA clone TTBA063B04 3',
 mRNA sequence.
 ACCESSION CR413015
 VERSION CR413015.1 GI:48681262
 KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 821)
 Croning, M.D.R., Ashurst, J.L., Taylor, R., Garrett, N. and Rogers, J.
 Sanger Xenopus tropicalis EST project 2001 (2004)
 Unpublished (2004)
 Contact: Croning MDR
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE_ID: TTBA063B04.q1kT7

This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.
Seq primer: T7.

FEATURES
Source Location/Qualifiers

1..821
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TDBA063b04"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_id="XGC-tailbud"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from tailbud. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 821;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
|||||
Db 269 AAAAAAAAAAGTTACAATTCAGAT 246

RESULT 48

LOCUS

CX329054 824 bp mRNA linear EST 04-JAN-2005
JGI_XZT67960.fwd NIH_XGC_troptads Xenopus tropicalis cDNA clone
IMAGE:7787016 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2004)

Other ESTs: JGI_XZT67960.rev

Contact: Lindquist, E.A., Richardson, P.

DOB Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of

California, Berkeley: http://tropicalis.berkeley.edu/home

cDNA Library Preparation: Richard M. Harland Laboratory, University

of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute; http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/LNL

http://image.lnl.gov

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone id and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Small insert: Based upon one or more sequencing reads of this clone

where vector sequence was present at both ends, this clone has been

determined to contain a cDNA insert on the order of 600-1000 bases.

Plate: XZT 0705 row: p column: 22

High quality sequence stop: 801.

Location/Qualifiers

1..824

/organism="Xenopus tropicalis"

/mol_type="mRNA"

/db_xref="taxon:8364"

/db_xref="taxon:8364"

/clone="IMAGE:7787016"
/cissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene
Electroten-Blue"

/clone_id="NIH_XGC_troptads"
/note="Vector: pCS108; Site_1: SalI; Site_2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS108 (http://mc.berkeley.edu/labs/harland/pages/plasmids.html)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 824;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
|||||
Db 623 AAAAAAAAAAGTTACAATTCAGAT 646

RESULT 49

LOCUS

CL101305 832 bp DNA linear GSS 05-JAN-2004
ISB1-38E19.Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-38E19,
genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2003)

Contact: Richard K Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Insert Length: 75000 Std Error: 0.00

Seq primer: Sp6 ATTAGTGACACTATAG

Class: BAC ends

High quality sequence start: 7

High quality sequence stop: 694.

Location/Qualifiers

1..832

/organism="Xenopus tropicalis"

/mol_type="genomic DNA"

/db_xref="taxon:8364"

/clone="ISB1-38E19"

/clone_id="ISB1"

/note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC

Library Segment 1"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 832;
Best Local Similarity 91.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
|||||
Db 120 AAAAAAAAAAGTTACAATTCAGAT 143

RESULT 50

CK315320/c

DEFINITION

ACCESSION

KEYWORDS

DOORCE
ORGANISM

PREFERENCE

AUTHORS

11

COMMENT

FEATURES
SOURCE

ORIGIN

Query Match	83.2%	Score 20.8	DB 7	Length 851
Best Local Similarity	91.7%	Pred. NO. 1.6e+03		
Matches	22	Conservative	2	Indels 0; Gaps 0;
QY	1	AAAAAAAAAGTCCCAATTCAGAT	24	
Db	39	AAAAAAAAAATCCCAATTCAAAT	16	

Search completed: December 14, 2005, 07:35:21
Job time : 1761.1 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:43:33 ; Search time 180.2 Seconds
(without alignments)
68.002 Million cell updates/sec

Title: US-10-681-773-7

Perfect score: 25
Sequence: 1 aaaaaaaaaagtcacatcagata 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4161359 seqs, 245077644 residues

Total number of hits satisfying chosen parameters: 8322718

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Published Applications NA New:
1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.4	77.6	235033	7 US-11-157-389-1	Sequence 1, Appl1
C 2	19.4	77.6	237326	7 US-11-157-389-2	Sequence 2, Appl1
C 3	18.8	75.2	1819	6 US-10-995-054A-35	Sequence 35, Appl1
C 4	18.8	75.2	59110	6 US-10-995-561-13324	Sequence 13324, A
C 5	18.6	74.4	201	6 US-10-995-561-74322	Sequence 74322, A
C 6	18.6	74.4	1252	6 US-10-750-185-62619	Sequence 62619, A
C 7	18.6	74.4	1543	6 US-10-750-185-53789	Sequence 53789, A
C 8	18.6	74.4	1953	6 US-10-750-185-53789	Sequence 53789, A
C 9	18.6	74.4	2118	6 US-10-750-185-53789	Sequence 53789, A
C 10	18.6	74.4	2655	6 US-10-750-185-53789	Sequence 53789, A
C 11	18.6	74.4	28277	6 US-10-750-185-53789	Sequence 53789, A
C 12	18.6	74.4	45517	6 US-10-995-561-13324	Sequence 13324, A
C 13	18.6	74.4	149419	7 US-11-112-908-49	Sequence 149419, A
C 14	18.6	74.4	166111	7 US-11-112-908-49	Sequence 149419, A
C 15	18.6	74.4	212716	7 US-11-121-086-95	Sequence 95, Appl1
C 16	18.6	74.4	398287	6 US-10-995-561-13324	Sequence 13324, A
C 17	18.2	72.8	833	7 US-11-112-908-173	Sequence 173, Appl1
C 18	18.2	72.8	1534	6 US-10-750-185-50355	Sequence 50355, A
C 19	18.2	72.8	2125	6 US-10-750-185-50355	Sequence 50355, A
C 20	18.2	72.8	2971	6 US-10-750-185-50355	Sequence 50355, A
C 21	18.2	72.8	150437	7 US-11-112-908-44	Sequence 44, Appl1
C 22	18.2	72.8	182314	7 US-11-112-908-45	Sequence 45, Appl1
C 23	18.2	72.8	203467	7 US-11-121-086-50	Sequence 50, Appl1

C 24	17.8	71.2	201	6 US-10-995-561-72216	Sequence 72216, A
C 25	17.8	71.2	1821	6 US-10-750-185-53499	Sequence 53499, A
C 26	17.8	71.2	2093	6 US-10-750-185-53499	Sequence 53499, A
C 27	17.8	71.2	2286	6 US-10-750-185-53499	Sequence 53499, A
C 28	17.8	71.2	3232	6 US-10-750-185-62480	Sequence 62480, A
C 29	17.8	71.2	47622	6 US-10-995-561-13343	Sequence 13343, A
C 30	17.8	71.2	67126	6 US-10-995-561-13342	Sequence 13342, A
C 31	17.8	71.2	340000	7 US-11-102-978-3	Sequence 3, Appl1
C 32	17.8	71.2	415117	6 US-10-995-561-13274	Sequence 13274, A
C 33	17.6	70.4	201	6 US-10-995-561-84829	Sequence 84829, A
C 34	17.6	70.4	600	6 US-10-750-185-2971	Sequence 2971, Appl1
C 35	17.6	70.4	600	6 US-10-750-185-19603	Sequence 19603, A
C 36	17.6	70.4	600	6 US-10-750-185-20932	Sequence 20932, A
C 37	17.6	70.4	600	6 US-10-750-185-45969	Sequence 45969, A
C 38	17.6	70.4	830	6 US-10-750-185-52225	Sequence 52225, A
C 39	17.6	70.4	838	6 US-10-750-185-54804	Sequence 54804, A
C 40	17.6	70.4	966	6 US-10-750-185-56618	Sequence 56618, A
C 41	17.6	70.4	1005	6 US-10-750-185-51843	Sequence 51843, A
C 42	17.6	70.4	1036	6 US-10-750-185-53216	Sequence 53216, A
C 43	17.6	70.4	1036	6 US-10-750-185-46890	Sequence 46890, A
C 44	17.6	70.4	1090	6 US-10-750-185-64207	Sequence 64207, A
C 45	17.6	70.4	1173	6 US-10-750-185-64186	Sequence 64186, A
C 46	17.6	70.4	1235	6 US-10-750-185-45149	Sequence 45149, A
C 47	17.6	70.4	1332	6 US-10-750-185-27505	Sequence 27505, A
C 48	17.6	70.4	1470	6 US-10-750-185-41796	Sequence 41796, A
C 49	17.6	70.4	1532	6 US-10-750-185-33207	Sequence 33207, A
C 50	17.6	70.4	1624	6 US-10-750-185-41678	Sequence 41678, A
C 51	17.6	70.4	2462	6 US-10-750-185-53879	Sequence 53879, A
C 52	17.6	70.4	60486	6 US-10-995-561-13330	Sequence 13310, A
C 53	17.6	70.4	96539	6 US-10-995-561-13289	Sequence 13289, A
C 54	17.6	70.4	150314	7 US-11-112-908-24	Sequence 24, Appl1
C 55	17.6	70.4	197096	7 US-11-121-086-107	Sequence 107, Appl1
C 56	17.6	70.4	380749	6 US-10-995-561-13216	Sequence 42775, A
C 57	17.6	70.4	4195235	6 US-10-995-561-44105	Sequence 44105, A
C 58	17.6	70.4	4195235	6 US-10-995-561-47062	Sequence 47062, A
C 59	17.2	68.8	201	6 US-10-995-561-47062	Sequence 47062, A
C 60	17.2	68.8	201	6 US-10-995-561-47062	Sequence 47062, A
C 61	17.2	68.8	201	6 US-10-995-561-47062	Sequence 47062, A
C 62	17.2	68.8	201	6 US-10-995-561-47062	Sequence 47062, A
C 63	17.2	68.8	201	6 US-10-995-561-47062	Sequence 47062, A
C 64	17.2	68.8	13432	6 US-10-750-185-55911	Sequence 55911, A
C 65	17.2	68.8	13432	6 US-10-750-185-55911	Sequence 55911, A
C 66	17.2	68.8	21848	6 US-10-995-561-13286	Sequence 13286, A
C 67	17.2	68.8	95050	6 US-10-995-561-13286	Sequence 13286, A
C 68	17.2	68.8	177623	7 US-11-112-908-41	Sequence 41, Appl1
C 69	17.2	68.8	189252	7 US-11-121-086-54	Sequence 54, Appl1
C 70	17.2	68.8	645179	6 US-10-995-561-13293	Sequence 13293, A
C 71	17.2	68.8	1125000	6 US-10-995-561-13286	Sequence 13286, A
C 72	17.2	68.8	201	6 US-10-995-561-22526	Sequence 22526, A
C 73	17.2	68.8	201	6 US-10-995-561-35477	Sequence 35477, A
C 74	17.2	68.8	201	6 US-10-995-561-46248	Sequence 46248, A
C 75	17.2	68.8	201	6 US-10-995-561-53820	Sequence 53820, A
C 76	17.2	68.8	201	6 US-10-995-561-53873	Sequence 53873, A
C 77	17.2	68.8	201	6 US-10-995-561-58868	Sequence 58868, A
C 78	17.2	68.8	201	6 US-10-995-561-58870	Sequence 58870, A
C 79	17.2	68.8	201	6 US-10-995-561-58872	Sequence 58872, A
C 80	17.2	68.8	201	6 US-10-995-561-59502	Sequence 59502, A
C 81	17.2	68.8	201	6 US-10-995-561-59537	Sequence 59537, A
C 82	17.2	68.8	201	6 US-10-995-561-69364	Sequence 69364, A
C 83	17.2	68.8	201	6 US-10-995-561-71237	Sequence 71237, A
C 84	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 85	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 86	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 87	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 88	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 89	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 90	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 91	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 92	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 93	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 94	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 95	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A
C 96	17.2	68.8	201	6 US-10-995-561-71266	Sequence 71266, A

97 17 68.0 1579 6 US-10-750-185-36262 Sequence 36262, A
C 98 17 68.0 1588 6 US-10-750-185-39457 Sequence 39457, A
C 99 17 68.0 1667 6 US-10-750-185-43357 Sequence 43357, A
C 100 17 68.0 1744 6 US-10-750-185-41171 Sequence 41171, A
C 101 17 68.0 1841 6 US-10-750-185-41033 Sequence 41033, A
C 102 17 68.0 1866 6 US-10-750-185-42514 Sequence 42514, A
C 103 17 68.0 2124 6 US-10-750-185-30610 Sequence 30610, A
C 104 17 68.0 2174 6 US-10-750-185-31302 Sequence 31302, A
C 105 17 68.0 2286 6 US-10-750-185-38924 Sequence 38924, A
C 106 17 68.0 2498 6 US-10-750-185-25482 Sequence 25482, A
C 107 17 68.0 2569 6 US-10-750-185-58501 Sequence 58501, A
C 108 17 68.0 2585 6 US-10-750-185-48577 Sequence 48577, A
C 109 17 68.0 2859 6 US-10-750-185-24500 Sequence 24500, A
C 110 17 68.0 2861 6 US-10-750-185-57413 Sequence 57413, A
C 111 17 68.0 2892 6 US-10-750-185-26550 Sequence 26550, A
C 112 17 68.0 3415 6 US-10-750-185-11178 Sequence 11178, A
C 113 17 68.0 3571 6 US-10-750-185-33894 Sequence 33894, A
C 114 17 68.0 4038 6 US-10-750-185-62503 Sequence 62503, A
C 115 17 68.0 4163 6 US-10-995-561-13497 Sequence 13497, A
C 116 17 68.0 27902 6 US-10-995-561-13462 Sequence 13462, A
C 117 17 68.0 32070 6 US-10-995-561-13317 Sequence 13317, A
C 118 17 68.0 45268 6 US-10-995-561-13203 Sequence 13203, A
C 119 17 68.0 46752 6 US-10-995-561-13410 Sequence 13410, A
C 120 17 68.0 65931 6 US-10-995-561-13254 Sequence 13254, A
C 121 17 68.0 80450 7 US-11-117-187-201 Sequence 201, App
C 122 17 68.0 86585 7 US-11-117-187-198 Sequence 198, App
C 123 17 68.0 103660 6 US-10-995-561-13253 Sequence 13253, A
C 124 17 68.0 134174 7 US-11-121-086-99 Sequence 99, App1
C 125 17 68.0 148935 6 US-10-995-561-13308 Sequence 13308, A
C 126 17 68.0 149382 6 US-10-995-561-13272 Sequence 13272, A
C 127 17 68.0 151169 7 US-11-121-086-38 Sequence 38, App1
C 128 17 68.0 155515 7 US-11-112-908-42 Sequence 42, App1
C 129 17 68.0 155989 7 US-11-121-086-57 Sequence 57, App1
C 130 17 68.0 161726 7 US-11-112-908-48 Sequence 48, App1
C 131 17 68.0 161726 7 US-11-112-908-52 Sequence 52, App1
C 132 17 68.0 170837 7 US-11-121-086-97 Sequence 97, App1
C 133 17 68.0 177175 7 US-11-121-086-79 Sequence 79, App1
C 134 17 68.0 177623 7 US-11-112-908-41 Sequence 41, App1
C 135 17 68.0 180654 7 US-11-121-086-58 Sequence 58, App1
C 136 17 68.0 193363 7 US-11-112-908-32 Sequence 32, App1
C 137 17 68.0 199321 7 US-11-121-086-10 Sequence 10, App1
C 138 17 68.0 235033 7 US-11-157-389-1 Sequence 1, App1
C 139 17 68.0 237326 7 US-11-157-389-2 Sequence 2, App1
C 140 17 68.0 317876 6 US-10-995-561-13227 Sequence 13227, A
C 141 17 68.0 380749 6 US-10-995-561-13216 Sequence 13216, A
C 142 17 68.0 403278 6 US-10-995-561-13421 Sequence 13421, A
C 143 17 68.0 403278 6 US-10-995-561-13421 Sequence 13421, A
C 144 17 68.0 645179 6 US-10-995-561-13293 Sequence 13293, A
C 145 17 68.0 1082144 7 US-11-117-187-211 Sequence 211, App
C 146 16.8 67.2 887 6 US-10-750-185-47186 Sequence 47186, A
C 147 16.8 67.2 1083 6 US-10-750-185-31151 Sequence 31151, A
C 148 16.8 67.2 1515 6 US-10-750-185-35828 Sequence 35828, A
C 149 16.8 67.2 1574 6 US-10-750-185-49423 Sequence 49423, A
C 150 16.8 67.2 1748 6 US-10-750-185-57303 Sequence 57303, A

ALIGNMENTS

RESULT 1
US-11-157-389-1/c
Sequence 1, Application US/1157389
Publication No. US20050266481A1
GENERAL INFORMATION:
APPLICANT: Ruddy, David A.
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: POLYMORPHISMS IN THE REGION OF THE HUMAN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York

STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/11/157,389
FILING DATE: 20-June-2005
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/852,495
FILING DATE: 07-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0057-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 235033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-11-157-389-1
Query Match 77.6%; Score 19.4; DB 7; Length 235033;
Best Local Similarity 95.2%; Pred. No. 62;
Matches 20; Conservative 0; Mismatches 1; Gaps 0;
Oy 1 AAAAAAAAAAGTCCATTCA 21
Db 187050 AAAAAAAAAAGTCCATTCA 187030
RESULT 2
US-11-157-389-2/c
Sequence 2, Application US/1157389
Publication No. US20050266481A1
GENERAL INFORMATION:
APPLICANT: Ruddy, David A.
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: POLYMORPHISMS IN THE REGION OF THE HUMAN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/11/157,389
FILING DATE: 20-June-2005
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/852,495
FILING DATE: 07-MAY-1997

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Polissant, Brian M
; REGISTRATION NUMBER: 28,462
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237326 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-11-157-389-2

Query Match          77.6%; Score 19.4; DB 7; Length 237326;
Best Local Similarity 95.2%; Pred. No: 62;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 21
Db 189274 AAAAAAAAAAGTCCCAATTCA 189254

RESULT 3
US-10-955-054A-35/c
; Sequence 35, Application US/10955054A
; Publication No. US20050266420A1
; GENERAL INFORMATION:
; APPLICANT: PUSZTAI, LAJOS
; APPLICANT: SYMMANS, W. FRASER
; APPLICANT: HESS, KENNETH R.
; APPLICANT: AYERS, MARK
; APPLICANT: STEC, JAMES
; TITLE OF INVENTION: MULTIGENE PREDICTORS OF RESPONSE TO CHEMOTHERAPY
; FILE REFERENCE: UTXC:880US
; CURRENT APPLICATION NUMBER: US/10/955,054A
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 1819
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-955-054A-35

Query Match          75.2%; Score 18.8; DB 6; Length 1819;
Best Local Similarity 90.9%; Pred. No: 62;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 22
Db 609 AAAAAAAAAAGTCCCAATTCA 588

RESULT 4
US-10-995-561-13324/c
; Sequence 13324, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13324
; LENGTH: 59110
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(59110)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13324

Query Match          75.2%; Score 18.8; DB 6; Length 59110;
Best Local Similarity 90.9%; Pred. No: 92;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGTCCCAATTGAGAT 24
Db 15944 AAAAAAAAAAGTCCCAATTGAGAT 15923

RESULT 5
US-10-995-561-74322
; Sequence 74322, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74322
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-74322

Query Match          74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No: 58;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTGAGATA 25
Db 12 AAAAAAAAAAGTCCCAATTGAGATA 36

RESULT 6
US-10-750-185-62619
; Sequence 62619, Application US/10750185
; Publication No. US2005026603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62619
; LENGTH: 1252
; TYPE: DNA
; ORGANISM: Bovine 19866881309218
```

US-10-750-185-62619

Query Match 74.4%; Score 18.6; DB 6; Length 1252;
Best Local Similarity 84.0%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 124 AAAAAAAAAATCCCAAAACACATA 148

RESULT 7

US-10-750-185-53789/C
; Sequence 53789, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53789
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Bovine 198668134851
US-10-750-185-53789

Query Match 74.4%; Score 18.6; DB 6; Length 1543;
Best Local Similarity 84.0%; Pred. No. 73;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 1311 AAAAAAAAAAGCCCAAGTTCAGAAA 1287

RESULT 8

US-10-750-185-36272
; Sequence 36272, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36272
; LENGTH: 1953
; TYPE: DNA
; ORGANISM: Bovine 19866880708703
US-10-750-185-36272

Query Match 74.4%; Score 18.6; DB 6; Length 1953;

Best Local Similarity 84.0%; Pred. No. 75;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 206 AAAAAAAAAAGTCTTAAACAGGTA 230

RESULT 9

US-10-750-185-49665
; Sequence 49665, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49665
; LENGTH: 2118
; TYPE: DNA
; ORGANISM: Bovine 19866880572776
US-10-750-185-49665

Query Match 74.4%; Score 18.6; DB 6; Length 2118;
Best Local Similarity 84.0%; Pred. No. 76;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 505 AAAAAAAAAAGTATCAGTTCAGTTA 529

RESULT 10

US-10-750-185-25238/C
; Sequence 25238, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25238
; LENGTH: 2655
; TYPE: DNA
; ORGANISM: Bovine 19866881085665
US-10-750-185-25238

Query Match 74.4%; Score 18.6; DB 6; Length 2655;
Best Local Similarity 84.0%; Pred. No. 78;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 716 AATAAATAAGTCTCAATTCAGATA 692

RESULT 11
US-10-995-561-13241
; Sequence 13241, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13241
; LENGTH: 28277
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(28277)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13241

Query Match 74.4%: Score 18.6; DB 6; Length 28277;
Best Local Similarity 84.0%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 1297 AAAAAAAAAAGCCCACTCAAGTA 1321

RESULT 12
US-10-995-561-13455
; Sequence 13455, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13455
; LENGTH: 45517
; TYPE: DNA
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-995-561-13455

Query Match 74.4%: Score 18.6; DB 6; Length 45517;
Best Local Similarity 84.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 20241 AAAAAAAAAATCCATTTGAGATA 20265

RESULT 13
US-11-112-908-49
; Sequence 49, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: HARRIS, Cole
```

```
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 49
; LENGTH: 149419
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-49

Query Match 74.4%: Score 18.6; DB 7; Length 149419;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 148267 AAAAAACAAGTCCCAATTTGAATA 148291

RESULT 14
US-11-112-908-47
; Sequence 47, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 47
; LENGTH: 166111
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-47

Query Match 74.4%: Score 18.6; DB 7; Length 166111;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 125367 AAAAAACAAGTCCCAATTTGAATA 125391

RESULT 15
US-11-121-086-95/C
; Sequence 95, Application US/11121086
; Publication No. US2005026459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
```

```
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/566,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO: 95
; LENGTH: 212716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-95

Query Match          74.4%; Score 18.6; DB 7; Length 212716;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 60083 AAACAAAAGGTACAAATTCAGATA 60059

RESULT 16
US-10-995-561-13396/C
; Sequence 13396, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13396
; LENGTH: 398287
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(398287)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-

Query Match          74.4%; Score 18.6; DB 6; Length 398287;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 229943 AAAAAAAAAAGTTAAATTCATATA 229919

RESULT 17
US-11-112-908-173/C
; Sequence 173, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: HARRIS, Cole
; APPLICANT: DAVIS, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
```

```
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 173
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-173

Query Match          72.8%; Score 18.2; DB 7; Length 833;
Best Local Similarity 87.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 147 AAAAAAAAAATGACCAATTCCTGA 125

RESULT 18
US-10-750-185-50355/C
; Sequence 50355, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50355
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-50355

Query Match          72.8%; Score 18.2; DB 6; Length 1534;
Best Local Similarity 87.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 365 AAAAAAAAAATCCAAATTCAGA 343

RESULT 19
US-10-750-185-54826/C
; Sequence 54826, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
```



```

; SOFTWARE: Patentin version 3.1
; SEQ ID NO 54826
; LENGTH: 2125
; TYPE: DNA
; ORGANISM: Bovine 1986680958781
US-10-750-185-54826

Query Match          72.8%; Score 18.2; DB 6; Length 2125;
Best Local Similarity 87.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 1650 AAAAAAAAAACTTACCAATTCAGA 1628

RESULT 20
US-10-750-185-61449
; Sequence 61449, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 61449
; LENGTH: 2971
; TYPE: DNA
; ORGANISM: Bovine 1986680852350
US-10-750-185-61449

Query Match          72.8%; Score 18.2; DB 6; Length 2971;
Best Local Similarity 87.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 2164 AAAAAAAAAATCCAAATGCAGAT 2186

RESULT 21
US-11-112-908-44/c
; Sequence 44, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 44
```

```

; LENGTH: 150437
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-44

Query Match          72.8%; Score 18.2; DB 7; Length 150437;
Best Local Similarity 87.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 26604 AAAAAAAAAATGACCAATTCGA 26582

RESULT 22
US-11-112-908-45/c
; Sequence 45, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIOR FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 45
; LENGTH: 182314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-45

Query Match          72.8%; Score 18.2; DB 7; Length 182314;
Best Local Similarity 87.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 89244 AAAAAAAAAATGACCAATTCGA 89222

RESULT 23
US-11-121-086-50/c
; Sequence 50, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 0918.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 50
; LENGTH: 203467
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-50

Query Match          72.8%; Score 18.2; DB 7; Length 203467;
Best Local Similarity 87.0%; Pred. No. 1.8e+02;
```

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTCCCAATCGA 23
Db 49343 AAAAAAAAAAGTCCCAATCGA 49321

RESULT 24
US-10-995-561-72216/c
; Sequence 72216, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: C001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72216
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-72216

Query Match 71.2%; Score 17.8; DB 6; Length 201;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATCGA 21
Db 197 AAAAAAAAAAGTCCCAATCGA 177

RESULT 25
US-10-750-185-53499
; Sequence 53499, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53499
; LENGTH: 1821
; TYPE: DNA
; ORGANISM: Bovine 19866881161328
US-10-750-185-53499

Query Match 71.2%; Score 17.8; DB 6; Length 1821;
Best Local Similarity 90.5%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATCGA 21
Db 533 AAAAAAAAAATCCCTATTCA 553

RESULT 26
US-10-750-185-36090

; Sequence 36090, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36090
; LENGTH: 2093
; TYPE: DNA
; ORGANISM: Bovine 19866880838363
US-10-750-185-36090

Query Match 71.2%; Score 17.8; DB 6; Length 2093;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATCGA 21
Db 1843 AAAAAAAAAAGTCCCAATCGA 1863

RESULT 27
US-10-750-185-34485
; Sequence 34485, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34485
; LENGTH: 2286
; TYPE: DNA
; ORGANISM: Bovine 19866880720505
US-10-750-185-34485

Query Match 71.2%; Score 17.8; DB 6; Length 2286;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATCGA 21
Db 2263 AAAAAAAAAAGTCCCAATCGA 2283

RESULT 28
US-10-750-185-62480
; Sequence 62480, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:

```

; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62480
; LENGTH: 3232
; TYPE: DNA
; ORGANISM: Bovine 19866880764359
US-10-750-185-62480

Query Match      71.2%; Score 17.8; DB 6; Length 3232;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      3 AAAAAAAAAATCCCAATTCAGA 23
DB      235 AAAAAAAAAATCCCAATTCAGA 255

RESULT 29
US-10-995-561-13431/C
; Sequence 13431, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13431
; LENGTH: 47622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13431

Query Match      71.2%; Score 17.8; DB 6; Length 47622;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAAAAAAAAATCCCAATTC A 21
DB      17723 AAAAAAAAAATCCCAACTCA 17703

RESULT 30
US-10-995-561-13342/C
; Sequence 13342, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 13342
; LENGTH: 67126
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13342

Query Match      71.2%; Score 17.8; DB 6; Length 67126;
Best Local Similarity 90.5%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAAAAAAAAATCCCAATTC A 21
DB      9349 AAAAAAAAAATCCCAAGTCA 9329

RESULT 31
US-11-102-978-3
; Sequence 3, Application US/11102978
; Publication No. US20050250142A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Technology Transfer Office
; APPLICANT: University of Utah Research Foundation
; TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
; FILE REFERENCE: 0274-5537.1US
; CURRENT APPLICATION NUMBER: US/11/102,978
; CURRENT FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: PCT/US2003/033152
; PRIOR FILING DATE: 2003-10-18
; PRIOR APPLICATION NUMBER: 60/419,576
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 340000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (56948)..(57115)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80006)..(81089)
; OTHER INFORMATION: Gene VDAC2P; voltage-dependent anion channel isoform 2 pseudogen
; FEATURE:
; NAME/KEY: exon
; LOCATION: (167308)..(167438)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: exon
; LOCATION: (216732)..(216833)
; OTHER INFORMATION: C21orf34 exon
US-11-102-978-3

Query Match      71.2%; Score 17.8; DB 7; Length 340000;
Best Local Similarity 90.5%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAAAAAAAAATCCCAATTC A 21
DB      291322 AAAAAAAAAATCAATTC A 291342

RESULT 32
US-10-995-561-13274/C
; Sequence 13274, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
```

```
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13274
/ LENGTH: 415117
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)_(415117)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13274
```

```
Query Match 71.2%; Score 17.8; DB 6; Length 415117;
Best Local Similarity 90.5%; Pred. No. 2,5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTCCCAATTCATCA 21
Db 98732 AAGAAAAAAGTCAATTCATCA 98712
```

```
RESULT 33
US-10-995-561-51999
/ Sequence 51999, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 51999
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-51999
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 1,4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 149 AAAAAAAAAAGTCAATTAAGTT 172
```

```
RESULT 34
US-10-995-561-84829/C
/ Sequence 84829, Application US/10995561
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 84829
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-84829
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 201;
```

```
Best Local Similarity 83.3%; Pred. No. 1,4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 166 AAAATCAAGTCCCAATTCAGAGA 143
```

```
RESULT 35
US-10-750-185-2971
/ Sequence 2971, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 2971
/ LENGTH: 600
/ TYPE: DNA
/ ORGANISM: Bovine MM8T08661
US-10-750-185-2971
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 600;
Best Local Similarity 83.3%; Pred. No. 1,6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 319 AAAAAAAAAATCCCAATTTAAAAA 342
```

```
RESULT 36
US-10-750-185-19603
/ Sequence 19603, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 19603
/ LENGTH: 600
/ TYPE: DNA
/ ORGANISM: Bovine MM8T02264
US-10-750-185-19603
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 600;
Best Local Similarity 83.3%; Pred. No. 1,6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 523 AAAAAAAAAAGTCCCAATTCATAT 546

RESULT 37

US-10-750-185-20932
; Sequence 20932, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20932
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT13642
US-10-750-185-20932

Query Match 70.4%; Score 17.6; DB 6; Length 600;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 192 AAAAAATACGACTCAATTCAGAT 215

RESULT 38

US-10-750-185-45969/C
; Sequence 45969, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45969
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Bovine 19866881044224
US-10-750-185-45969

Query Match 70.4%; Score 17.6; DB 6; Length 714;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 711 AAGAAAAAGACCAATTCAGAT 688

RESULT 39
US-10-750-185-52225

; Sequence 52225, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52225
; LENGTH: 830
; TYPE: DNA
; ORGANISM: Bovine 19866881193001
US-10-750-185-52225

Query Match 70.4%; Score 17.6; DB 6; Length 830;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 299 AAAAAATACGACTCAATTCAGAT 322

RESULT 40

US-10-750-185-54804
; Sequence 54804, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54804
; LENGTH: 838
; TYPE: DNA
; ORGANISM: Bovine 19866881191907
US-10-750-185-54804

Query Match 70.4%; Score 17.6; DB 6; Length 838;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 556 AAAAAAAAAAGTACCAATTCGAAT 579

RESULT 41

```
US-10-750-185-56618
; Sequence 56618, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56618
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Bovine 19866880682112
US-10-750-185-56618

Query Match      70.4%; Score 17.6; DB 6; Length 966;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 7 AAAAAAAAAATAAACCATAATTAAGAT 30

RESULT 42
US-10-750-185-51843/c
; Sequence 51843, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51843
; LENGTH: 1005
; TYPE: DNA
; ORGANISM: Bovine 19866881513039
US-10-750-185-51843

Query Match      70.4%; Score 17.6; DB 6; Length 1005;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 770 AAAAAAAAAATTCCAATCAGAT 747

RESULT 43
US-10-750-185-53216/c
; Sequence 53216, Application US/10750185
; Publication No. US20050260603A1
```

```
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53216
; LENGTH: 1036
; TYPE: DNA
; ORGANISM: Bovine 19866881289993
US-10-750-185-53216

Query Match      70.4%; Score 17.6; DB 6; Length 1036;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 499 AAAAAAAAAATCACCATAATTCAGTA 476

RESULT 44
US-10-750-185-46890
; Sequence 46890, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46890
; LENGTH: 1045
; TYPE: DNA
; ORGANISM: Bovine 19866881306604
US-10-750-185-46890

Query Match      70.4%; Score 17.6; DB 6; Length 1045;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 14 AAAAAAAAAATGACCTAATTCAAAT 37

RESULT 45
US-10-750-185-64207/c
; Sequence 64207, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
```

```

; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64207
; LENGTH: 1090
; TYPE: DNA
; ORGANISM: Bovine 1986681234908
US-10-750-185-64207

```

```

Query Match          70.4% Score 17.6; DB 6; Length 1090;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 794 AAAAAAAAAAACCAATTCAGCA 771

```

```

RESULT 46
US-10-750-185-45149/c
; Sequence 45149, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45149
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Bovine 19866881021600
US-10-750-185-45149

```

```

Query Match          70.4% Score 17.6; DB 6; Length 1173;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 2 AAAAAAAAAAGTCCCAATTCAGATA 25
Db 964 AAAAAAAAAAGTCCCAATTCAGTA 941

```

```

RESULT 47
US-10-750-185-64186/c
; Sequence 64186, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom

```

```

; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64186
; LENGTH: 1235
; TYPE: DNA
; ORGANISM: Bovine 19866880339909
US-10-750-185-64186

```

```

Query Match          70.4% Score 17.6; DB 6; Length 1235;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 1040 AAAAAAAAAATCCCAATCTGTT 1017

```

```

RESULT 48
US-10-750-185-27505
; Sequence 27505, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27505
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Bovine 19866880569471
US-10-750-185-27505

```

```

Query Match          70.4% Score 17.6; DB 6; Length 1332;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAAGTCCCAATTCAGAT 24
Db 1097 AAAAAAAAAATCCCAATCATAT 1120

```

```

RESULT 49
US-10-750-185-41796
; Sequence 41796, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

```

```

; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41796
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Bovine 19866881268786
US-10-750-185-41796

```

```

Query Match      70.4%; Score 17.6; DB 6; Length 1470;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAGTCCCAATTCAGAT 24
          |||||
Db      244 AAAAAATAAATCCCAATCCACAT 267

```

RESULT 50

```

US-10-750-185-33207
; Sequence 33207, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 33207
; LENGTH: 1532
; TYPE: DNA
; ORGANISM: Bovine 19866880946041
US-10-750-185-33207

```

```

Query Match      70.4%; Score 17.6; DB 6; Length 1532;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAAGTCCCAATTCAGAT 24
          |||||
Db      248 ACAATTAAGACCAACTCAGAT 271

```

```

Search completed: December 14, 2005, 11:40:45
Job time : 188.2 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-8

Perfect score: 25
Sequence: 1 aaaaaaaaaagttcccaactcagata 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5683141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database:

GenEmbl:
1: gb_ha:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pt:*
9: gb_to:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vt:*
14: gb_ncg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	25	100.0	88814 8 AC010270	AC010270 Homo sapi
2	25	100.0	103701 8 AC008434	AC008434 Homo sapi
3	25	100.0	163303 14 AC104467	AC104467 Homo sapi
4	25	100.0	223885 8 AV275681	AV275681 Homo sapi
5	23.4	93.6	140712 14 AC002223	AC002223 Homo sapi
6	22.4	89.6	174224 14 AC148789	AC148789 Homo sapi
7	21.8	87.2	639 10 BV374610	BV374610 S231P614
8	21.8	87.2	777 10 BV624515	BV624515 S215P604
9	21.8	87.2	66834 14 AC141798	AC141798 Apis mell
10	21.8	87.2	125051 8 AC005152	AC005152 Human DNA
11	21.8	87.2	126150 8 AC005152	AC005152 Human DNA
12	21.8	87.2	166882 8 AC084783	AC084783 Homo sapi
13	21.8	87.2	188881 8 AC157526	AC157526 Homo sapi
14	21.8	87.2	189278 14 AC055787	AC055787 Homo sapi
15	21.8	87.2	199808 9 AC103600	AC103600 Mus muscu
16	21.8	87.2	220832 14 AC164224	AC164224 Bos tauru
17	21.4	85.6	110000 8 AB128049_15	Continuation (16 o
18	21.4	85.6	147539 8 AC148704	AC148704 Macaca mu

19	21.4	85.6	158818 8 AC148710	AC148710 Macaca mu
20	21.4	85.6	163279 8 AC073263	AC073263 Homo sapi
21	21.4	85.6	186491 8 AC148705	AC148705 Macaca mu
22	21	84.0	209056 14 AC159797	AC159797 Bos tauru
23	20.8	83.2	307 8 HS214XG7	HS214XG7
24	20.8	83.2	421 10 BV003898	BV003898
25	20.8	83.2	4264 1 AB183841	AB183841 Streptoco
26	20.8	83.2	13356 1 AB011418	AB011418 Leptospir
27	20.8	83.2	31271 6 C0609503	C0609503 Sequence
28	20.8	83.2	58041 9 BX510318	BX510318 Mouse DNA
29	20.8	83.2	62860 8 AL162395	AL162395 Human DNA
30	20.8	83.2	96174 8 AP000243	AP000243 Homo sapi
31	20.8	83.2	100000 8 AP000203	AP000203 Homo sapi
32	20.8	83.2	106659 14 AC020331	AC020331 Drosophi1
33	20.8	83.2	110000 14 AC153420	AC153420 Mus muscu
34	20.8	83.2	110000 14 AC162929_1	Continuation (2 of
35	20.8	83.2	110000 14 BX640582_1	Continuation (2 of
36	20.8	83.2	110000 15 AP008214_099	Continuation (100
37	20.8	83.2	120562 15 AC124958	AC124958 Medicago
38	20.8	83.2	142202 15 AC144474	AC144474 Medicago
39	20.8	83.2	144833 14 AC063921	AC063921 Homo sapi
40	20.8	83.2	145633 14 BS000107	BS000107 Pan trogl
41	20.8	83.2	149800 14 AC021981	AC021981 Homo sapi
42	20.8	83.2	153337 15 CFE828413	CFE828413
43	20.8	83.2	155456 14 AC027753	AC027753 Homo sapi
44	20.8	83.2	160732 8 AC018647	AC018647 Homo sapi
45	20.8	83.2	161172 2 AC007839	AC007839 Drosophi1
46	20.8	83.2	162020 14 BX950208	BX950208 Danio rer
47	20.8	83.2	166423 15 AP005697	AP005697 Oryza sat
48	20.8	83.2	170030 8 AC005829	AC005829 Homo sapi
49	20.8	83.2	172546 14 AC024604	AC024604 Homo sapi
50	20.8	83.2	174420 14 AC166110	AC166110 Mus muscu
51	20.8	83.2	174662 2 CEY59A8B	CEY59A8B Caenorhab
52	20.8	83.2	178169 8 AP002765	AP002765 Homo sapi
53	20.8	83.2	178431 14 AP002749	AP002749 Homo sapi
54	20.8	83.2	179282 14 AC147797	AC147797 Mus muscu
55	20.8	83.2	179564 8 AC136677	AC136677 Homo sapi
56	20.8	83.2	179954 8 AC007271	AC007271 Homo sapi
57	20.8	83.2	181471 8 AC069539	AC069539 Homo sapi
58	20.8	83.2	187685 8 AC127032	AC127032 Homo sapi
59	20.8	83.2	190265 14 AC126604	AC126604 Homo sapi
60	20.8	83.2	190844 14 AC073990	AC073990 Homo sapi
61	20.8	83.2	192514 14 AC162014	AC162014 Bos tauru
62	20.8	83.2	195876 9 AC124133	AC124133 Mus muscu
63	20.8	83.2	200157 9 AC140409	AC140409 Mus muscu
64	20.8	83.2	223070 9 AC118546	AC118546 Mus muscu
65	20.8	83.2	224180 14 BX649286	BX649286 Danio rer
66	20.8	83.2	227395 9 AC119248	AC119248 Mus muscu
67	20.8	83.2	230105 5 BX537358	BX537358 Zebrafish
68	20.8	83.2	255997 14 CR759880	CR759880 Danio rer
69	20.8	83.2	268901 14 AC109139	AC109139 Mus muscu
70	20.8	83.2	282983 14 AC161916	AC161916 Bos tauru
71	20.8	83.2	293024 14 CEY59A8	CEY59A8 Caenorhabd1
72	20.8	83.2	306854 2 AE003797	AE003797 Drosophi1
73	20.8	83.2	317781 14 AC163164	AC163164 Bos tauru
74	20.8	83.2	340000 9 AP001709	AP001709 Homo sapi
75	20.4	81.6	142481 8 AL645976	AL645976 Mouse DNA
76	20.4	81.6	137547 14 AC140684	AC140684 Rattus no
77	20.4	81.6	141655 5 BX276108	BX276108 Zebrafish
78	20.4	81.6	146370 9 AC145550	AC145550 Mus muscu
79	20.4	81.6	148337 14 AC021509	AC021509 Homo sapi
80	20.4	81.6	152017 14 AC157838	AC157838 Oryzomys
81	20.4	81.6	158840 8 AC068812	AC068812 Homo sapi
82	20.4	81.6	164622 9 AC079436	AC079436 Rattus no
83	20.4	81.6	199710 5 CR391920	CR391920 Zebrafish
84	20.4	81.6	201794 8 CENS0180W	AL109750 Human chr
85	20.4	81.6	214546 9 AC164165	AC164165 Mus muscu
86	20.4	81.6	232195 14 AC106493	AC106493 Rattus no
87	20.2	80.8	201 10 BV175580	BV175580 sqm81517
88	20.2	80.8	313 6 AR525984	AR525984 Sequence
89	20.2	80.8	625 10 BV050759	BV050759 S208P606
90	20.2	80.8	721 10 BV631984	BV631984 S215P6088
91	20.2	80.8	995 6 AR509755	AR509755 Sequence

C 92	20.2	80.8	1445	2	AF151385	AF151385	Cochliomy
C 93	20.2	80.8	1449	2	AF151384	AF151384	Cochliomy
C 94	20.2	80.8	3759	6	CO582935	CO582935	Sequence
C 95	20.2	80.8	3759	6	AX254461	AX254461	Sequence
C 96	20.2	80.8	4169	6	CO594518	CO594518	Sequence
C 97	20.2	80.8	5933	6	AX348364	AX348364	Sequence
C 98	20.2	80.8	5933	6	AX348365	AX348365	Sequence
C 99	20.2	80.8	6486	6	AX458534	AX458534	Sequence
C 100	20.2	80.8	9510	6	AX251192	AX251192	Sequence
C 101	20.2	80.8	9510	6	AX281452	AX281452	Sequence
C 102	20.2	80.8	10151	6	AX347334	AX347334	Sequence
C 103	20.2	80.8	10151	6	AX347335	AX347335	Sequence
C 104	20.2	80.8	27128	9	AF210403	AF210403	Mus muscu
C 105	20.2	80.8	27128	9	AF202978	AF202978	Caenorhab
C 106	20.2	80.8	39706	8	AC000073	AC000073	Homo sapi
C 107	20.2	80.8	43577	8	AC090670	AC090670	Homo sapi
C 108	20.2	80.8	49162	14	AC160518	AC160518	Strongylo
C 109	20.2	80.8	69230	8	AC078865	AC078865	Homo sapi
C 110	20.2	80.8	71132	5	AC099321	AC099321	Danio rer
C 111	20.2	80.8	80615	14	AC130491	AC130491	Homo sapi
C 112	20.2	80.8	83825	14	AC144343	AC144343	Medicago
C 113	20.2	80.8	84112	8	AC016255	AC016255	Homo sapi
C 114	20.2	80.8	84961	14	AC014810	AC014810	Drosophila
C 115	20.2	80.8	89034	14	AC165978	AC165978	Bos tauru
C 116	20.2	80.8	89528	14	AP007883	AP007883	Lotus cor
C 117	20.2	80.8	89528	14	AC007067	AC007067	Genomic s
C 118	20.2	80.8	99291	15	ATFM19	AL138643	Arabidops
C 119	20.2	80.8	99416	15	AC009398	AC009398	Genomic s
C 120	20.2	80.8	100152	8	AC000085	AC000085	Homo sapi
C 121	20.2	80.8	102873	15	ATP6635	AL353814	Arabidops
C 122	20.2	80.8	104632	8	AL136124	AL136124	Human DNA
C 123	20.2	80.8	106540	15	AC129091	AC129091	Medicago
C 124	20.2	80.8	107839	15	AC157507	AC157507	Medicago
C 125	20.2	80.8	108394	15	AC148971	AC148971	Medicago
C 126	20.2	80.8	108759	1	CP000084	Continuation (13 o	
C 127	20.2	80.8	109745	8	AC025469	AC025469	Homo sapi
C 128	20.2	80.8	110000	1	BA000017	Continuation (16 o	
C 129	20.2	80.8	110000	1	BA000018	Continuation (15 o	
C 130	20.2	80.8	110000	1	CP000084_04	Continuation (5 of	
C 131	20.2	80.8	110000	15	AP008082_11_058	Continuation (59 o	
C 132	20.2	80.8	110000	15	AP008821_11_059	Continuation (60 o	
C 133	20.2	80.8	110000	15	AP008821_11_129	Continuation (130	
C 134	20.2	80.8	111967	4	AC097587	AC097587	Sus scrofa
C 135	20.2	80.8	111935	15	AC135427	AC135427	Oryza sat
C 136	20.2	80.8	114956	8	AC002072	AC002072	Human PAC
C 137	20.2	80.8	117937	14	AL139146	AL139146	Homo sapi
C 138	20.2	80.8	125561	15	AC152499	AC152499	Medicago
C 139	20.2	80.8	126437	15	AC148918	AC148918	Medicago
C 140	20.2	80.8	127769	8	AC009088	AC009088	Homo sapi
C 141	20.2	80.8	133769	8	AC010362	AC010362	Homo sapi
C 142	20.2	80.8	133769	8	AC010429	AC010429	Homo sapi
C 143	20.2	80.8	137726	8	AC105919	AC105919	Homo sapi
C 144	20.2	80.8	138495	14	AC146931	AC146931	Sus scrofa
C 145	20.2	80.8	140127	14	AC141742	AC141742	Apis mell
C 146	20.2	80.8	141171	15	AC132493	AC132493	Oryza sat
C 147	20.2	80.8	141383	15	AC130726	AC130726	Oryza sat
C 148	20.2	80.8	144223	8	AC062015	AC062015	Homo sapi
C 149	20.2	80.8	145417	14	AC138169	AC138169	Sus scrofa
C 150	20.2	80.8	145676	15	AC135922	AC135922	Oryza sat

ALIGNMENTS

RESULT 1
AC010270/c AC010270 88814 bp DNA linear PRI 28-JAN-2003
DEFINITION Homo sapiens chromosome 5 clone CTC-491N17, complete sequence.
AC010270
AC010270.8 GI:2723611
KEYWORDS
HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 88814)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
JOURNAL
2 (bases 1 to 88814)
DOE Joint Genome Institute.
Direct Submission
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 88814)
DOE Joint Genome Institute.
Direct Submission
Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 88814)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (01-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
5 (bases 1 to 88814)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (28-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jan 28, 2003 this sequence version replaced gi:21637455.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 27.1.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 171,5kb). It is clipped at the overlaps with AC022493
and AC008434. The number of bases overlapped with AC022493 is 11961
and with AC008434 is 6448.
Location/Qualifiers
1..88814
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-491N17"

ORIGIN
Query Match 100.0%; Score 25; DB 8; Length 88814;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1
AAAAAAAAAGTTCATTCAGATA 25
|||||
AAAAAAAAAGTTCATTCAGATA 85853

RESULT 2
AC008434/c AC008434 103701 bp DNA linear PRI 01-OCT-2002
DEFINITION Homo sapiens chromosome 5 clone CTC-325J23, complete sequence.
AC008434
AC008434.5 GI:23396211
KEYWORDS
HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 103701)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
JOURNAL
2 (bases 1 to 103701)

FEATURES	source	misc_feature	ORIGIN
REFERENCE TITLE JOURNAL AUTHORS	DOE Joint Genome Institute. Direct Submission Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA 3 (bases 1 to 103701) DOE Joint Genome Institute and Stanford Human Genome Center.		
COMMENT	Submitted (01-OCT-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA On Oct 1, 2002 this sequence version replaced gi:13699352. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov		
FEATURES	Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu Quality: Phrap Quality >=40 99.5% of Sequence; Estimated Total Number of Errors is 0.4. NOTE: Shatter Libraries failed to resolve the dinucleotide repeat region from 60278 to 61126. Forced join at 60885. NOTE: This insert is not the entire sequence of the clone (entire sequence is 136kd). It is clipped at the overlap with AC018764. The number of bases overlapped is 17570.		
source	1..103701 /organism="Homo sapiens" /mol_type="genomic DNA" /dd_xref="taxon:9606" /chromosome="5" /clone="CFC-325J23" 60278..61126 /note="NOTE: Shatter Libraries failed to resolve the dinucleotide repeat region from 60278 to 61126. Forced join at 60885."		
misc_feature			
Query Match	100.0%; Score 25; DB 8; Length 103701;		
Best Local Similarity	100.0%; Pred. No. 24;		
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 AAAAAAAAAAGTCCATTCAGATA 25 		
Db	3523 AAAAAAAAAAGTCCATTCAGATA 3499		
RESULT 3 AC104467/c LOCUS DEFINITION ACCESSION VERSION SOURCE ORGANISM	AC104467 163303 bp DNA linear HTG 26-PEB-2002 Homo sapiens chromosome 3q clone RP11-21E14, WORKING DRAFT SEQUENCE, 4 unordered pieces. AC104467 AC104467.3 GI:18874187 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. Homo sapiens (human) Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 163303) Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C., Albrooks S.L., Amaratunga H.C., Are J.R., Ayale M., Banks T., Barbata J., Benton J., Blinage K., Blankenburg K., Bonnin D., Bouck J., Bowle S., Brileva M., Brown E., Brown M., Bryant N.P., Bunyah C., Burket P., Burkett C., Butrell K.L., Byrd N.C., Caron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C., Davy Carroll L., Dedertich D.A., Delaney K.R., Delgado O., Dem A.L., Ding Y., Dinh H.H., Donahewite K.J., Draper H., Dugan-Rocha S., Durbin K.J., Earhart C., Edgar D., Edwards C.C., Elhaj C., Escoto M., Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Franz P., Gabel A., Gao J., Garcia A., Garner T., Garza N., Gill R., Gorrell J. H., Guevara M., Gunaratne P., Hale S., Hamilton K., Harris C., Harris K., Hart M., Havila P., Hawes A., Hernandez J.,		

Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
 Homai, F., Howard, S., Huber, J., Huiyk, S., Hune, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Katsinos, E., Kelly, S., Khan, U., King, L., Kovah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W., Loubege, H.,
 Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Mathew, M., Mappa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Mitter, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okunnu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swatek, A., Taber, P., Tameiisa, A., Tameiisa, K., Tang, H.,
 Tansley, C., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Umanai, K., Vazquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlingon, S.,
 Williams, G., Williamson, A., Wlaczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
 Weinscock, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 163303)
 Worley, K.C.
 Direct Submission
 Submitted (12-DEC-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 163303)
 Worley, K.C.
 Direct Submission
 Submitted (26-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 24, 2002 this sequence version replaced gi:17933806.
 ----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: HCVm
 Center clone name: RP11-21E14
 ----- Summary Statistics -----
 Sequencing vector: Plasmid; M7789
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 162741 bases at least Q40
 Consensus quality: 163445 bases at least Q30
 Consensus quality: 163933 bases at least Q20
 Estimated insert size: 162788; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 7.2x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 102608: contig of 102608 bp in length
 * 102609 102708: gap of unknown length
 * 102709 143726: contig of 40018 bp in length
 * 142727 142826: gap of unknown length
 * 142827 160594: contig of 17768 bp in length

FEATURES	*	160595	160694: gap of unknown length
	*	160695	163303: contig of 2609 bp in length.
source		1.	163303
			/organism="Homo sapiens"
			/mol_type="genomic DNA"
			/db_xref="taxon:9606"
			/chromosome="3q"
			/clone="RP11-21E14"
gap		102509..102708	
		/estimated_length=unknown	
gap		142727..142826	
		/estimated_length=unknown	
gap		160595..160694	
		/estimated_length=unknown	
ORIGIN			
Query Match		100.0%	Score 25; DB 14; Length 163303;
Best Local Similarity		100.0%;	Pred. No. 21;
Matches	25; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
Qy	1	AAAAAAAAAGTTCATTTCAGATA	25
Db	137304	AAAAAAAAAGTTCATTTCAGATA	137280
RESULT 4			
LOCUS	AY275681	223885 bp	DNA linear PRI 23-APR-2003
DEFINITION	Homo sapiens mucs homolog 3 (E. coli) (MSH3) gene, complete cds.		
ACCESSION	AY275681		
VERSION	AY275681.1	GI:30089005	
KEYWORDS	.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 223885)		
AUTHORS	Rieder,M.-J., Livingston,R.-J., Daniels,M.-R., Chung,M.-W., Miyamoto,K.-E., Nguyen,C.-P., Nguyen,D.-A., Poel,C.L., Robertson,P.-D., Schackwitz,M.-S., Sherwood,J.-K., Witrak,L.A. and Nickerson,D.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-APR-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA		
COMMENT	To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://esg.gs.washington.edu).		
FEATURES			
source	1..223885		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
variation	175		
	/frequency="0.01"		
	/replace="c"		
variation	338		
	/frequency="0.01"		
	/replace="a"		
variation	775		
	/frequency="0.14"		
	/replace="c"		
variation	869		
	/frequency="0.12"		
	/replace="c"		
variation	880		
	/frequency="0.12"		
	/replace="c"		
variation	884		
	/frequency="0.47"		
	/replace="g"		
gene	<901..222770		
	/gene="MSH3"		

[illegible]

```

/ gene="MSH3"
/ frequency="0.22"
/ replace="g"
1872
/ gene="MSH3"
/ frequency="0.20"
/ replace="c"
1877
/ gene="MSH3"
/ frequency="0.41"
/ replace="t"
2089
/ gene="MSH3"
/ frequency="0.01"
/ replace="c"
2190
/ gene="MSH3"
/ frequency="0.02"
/ replace="g"
2415
/ gene="MSH3"
/ frequency="0.40"
/ replace="g"
2535
/ gene="MSH3"
/ frequency="0.23"
/ replace="a"
2546
/ gene="MSH3"
/ frequency="0.23"
/ replace="g"
2771
/ gene="MSH3"
/ frequency="0.22"
/ replace="c"
2807
/ gene="MSH3"
/ frequency="0.05"
/ replace="a"
2807
/ gene="MSH3"
/ frequency="0.22"
/ replace="c"
2807
/ gene="MSH3"
/ frequency="0.73"
/ replace="g"
2922
/ gene="MSH3"
/ frequency="0.20"
/ replace="g"
3208
/ gene="MSH3"
/ frequency="0.21"
/ replace="t"
3375
/ rpt_family="Alu"
/ rpt_type=dispersed
3589
/ gene="MSH3"
/ frequency="0.01"
/ replace="a"
3697
/ gene="MSH3"
/ note="Region not scanned for variation"
4252
/ rpt_family="Alu"
/ rpt_type=dispersed
4823
/ rpt_family="Alu"
/ rpt_type=dispersed
5141
/ rpt_family="L1"

```

```

repeat_region      /rpt_type=dispersed
                    5178..5474
                    /rpt_family="Alu"
                    /rpt_type=dispersed
repeat_region      5475..5830
                    /rpt_family="L1"
                    /rpt_type=dispersed
repeat_region      5832..6081
                    /rpt_family="L1"

Query Match      100.0%; Score 25; DB 8; Length 223885;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAAAAAAAAAGTTCCATTCAGATA 25
Db      3869 AAAAAAAAAAGTTCCATTCAGATA 3893

RESULT 5
AC022223
LOCUS      Homo sapiens chromosome 5 clone RP11-538B23, WORKING DRAFT
DEFINITION
ACCESSION      AC022223.18 GI:20335453
VERSION      HTG; HTGS PHASE1; HTGS _DRAFT.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 140712)
REFERENCE
AUTHORS      Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrooke,S.L., Amaratunga,H.C., Are,U.R., Ayale,M., Banks,T.,
Barbataia,J., Benton,J., Bimage,K., Blankenhuy,K., Bonnin,D.,
Bouck,J., Bowls,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dahorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsel,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtenarge,O., Lien,C., Liu,J., Liu,W., Louisge,H.,
Lozad,R.J., Lu,X., Lucier,A., Luchter,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N.,
Nguyen,N., Nickerson,B., Nwokenkwo,S., Ogih,M., Okunomi,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiller,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshari,N., Slason,I.,
Sodergren,E., Sonatke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vazquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
Williams,G., Williamson,A., Wlezyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 140712)
AUTHORS Morley, K.C.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 140712)
AUTHORS Morley, K.C.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Apr 28, 2002 this sequence version replaced gi:11128048.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAO
Center clone name: RP11-538B23
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-Primer Bodipy: 1% of reads
Chemistry: Dye-terminator Big Dye: 8% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 123715 bases at least Q40
Consensus quality: 130186 bases at least Q30
Consensus quality: 133609 bases at least Q20
Estimated insert size: 138821; sum-of-coverage estimation
Quality coverage: 4x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2466: contig of 2466 bp in length
* 2467 2566: gap of unknown length
* 2567 2514: contig of 248 bp in length
* 5215 5314: gap of unknown length
* 5315 8344: contig of 3030 bp in length
* 8345 8444: gap of unknown length
* 8445 12178: contig of 3734 bp in length
* 12179 12278: gap of unknown length
* 12279 17378: contig of 5100 bp in length
* 17379 17478: gap of unknown length
* 17479 21216: contig of 3738 bp in length
* 21217 25247: contig of 3931 bp in length
* 25248 25347: gap of unknown length
* 25348 31074: contig of 5727 bp in length
* 31075 31174: gap of unknown length
* 31175 37463: contig of 6289 bp in length
* 37464 37563: gap of unknown length
* 37564 44703: contig of 7140 bp in length
* 44704 44803: gap of unknown length
* 44804 51956: contig of 7153 bp in length
* 51957 52056: gap of unknown length
* 52057 58167: contig of 6111 bp in length
* 58168 58267: gap of unknown length
* 58268 65527: contig of 7260 bp in length
* 65528 73849: gap of unknown length
* 73850 73949: contig of 8222 bp in length
* 73950 81898: gap of unknown length
* 81898: contig of 7949 bp in length

FEATURES
source
* 81899 81998: gap of unknown length
* 81999 92965: contig of 10967 bp in length
* 92966 93065: gap of unknown length
* 93066 102625: contig of 9560 bp in length
* 102626 102725: gap of unknown length
* 102726 119048: contig of 16323 bp in length
* 119049 140712: gap of unknown length
* 119149 140712: contig of 21564 bp in length.
Location/Qualifiers
1. .140712
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-538B23"
2467. .2566
/estimated_length=unknown
5215. .5314
/estimated_length=unknown
8345. .8444
/estimated_length=unknown
12179. .12278
/estimated_length=unknown
17379. .17478
/estimated_length=unknown
21217. .21316
/estimated_length=unknown
25248. .25347
/estimated_length=unknown
31075. .31174
/estimated_length=unknown
37464. .37563
/estimated_length=unknown
44704. .44803
/estimated_length=unknown
51957. .52056
/estimated_length=unknown
58168. .58267
/estimated_length=unknown
65528. .65627
/estimated_length=unknown
73850. .73949
/estimated_length=unknown
81899. .81998
/estimated_length=unknown
92966. .93065
/estimated_length=unknown
102626. .102725
/estimated_length=unknown
119049. .119148
/estimated_length=unknown

ORIGIN
Query Match 93.6%; Score 23.4; DB 14; Length 140712;
Best Local Similarity 96.0%; Pred. No. 77;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 AAAAAAAAAATTCATTCAGATA 25
Db 123229 AAAAAAAAAAGTCCCATTCAGATA 123253

RESULT 6
AC148789
LOCUS
DEFINITION
Ocolemur garnettii clone CH256-28D19, WORKING DRAFT SEQUENCE, 7
ordered pieces.
AC148789
AC148789.2 GI:46849631
VERSION
HTG: HTGS PHASE2; HTGS DRAFT.
KEYWORDS
Ocolemur garnettii (small-eared galago)
Ocolemur garnettii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

TITLE The genome sequence of Canis familiaris
JOURNAL Unpublished (2004)
COMMENT

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA

Tel: 6172521477
Fax: 6172580903
Email: kersti@genome.wi.mit.edu

Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 639
Protocol:

WGS-discovery (WGS) :
Paired-end low-coverage whole genome shotgun reads were generated
from 9 breeds

(German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
Retriever, English
Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
Water Dog -100,000 each)
and five other canids (Chinese, Alaskan, Indian and Spanish Gray
Wolf as well as the
Californian Coyote).

The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
and SNP detection was
carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
485941 SNPs were
annotated with alleles from the boxer and the breed or canid from
which the particular
read came. The validation rate for these SNPs was estimated at
approximately 98%.

MGA-discovery (MGA) of Boxer/Poodle SNPs:
A second set of SNPs was generated using a similar methodology
except that the contigs
from the 1.5x poodle assembly (Kirkness 2003) were used instead of
WGS reads. Since this
sequence lacked base quality scores, arbitrary quality scores of
phred 40 were assigned
before the poodle sequence was placed uniquely on the CanFam1.0
boxer assembly and SNP
detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
with alleles from the
boxer and the poodle. The validation rate for these SNPs was
estimated at approximately 78%.

Internal-MGA-discovery (I-MGA) :
A third set of SNPs were discovered by comparing reads in the MGA
assembly. SNPs were
defined as mismatch positions that had a base quality of >= 30 on
both reads in a region
that aligned without gaps, and with at most one additional mismatch
in the ten flanking
bases. For each allele, at least one additional read had to confirm
it. 731476 SNPs were
annotated with alleles between the two boxer alleles. The
validation rate for these SNPs
was estimated at approximately 78%.

Location/Qualifiers
1. 639
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="BedlingtonTerrier"
/db_xref="taxon:9615"
/map=" + 3 22-568 84888057-84888604"
/clone_lib="BedlingtonTerrier"
<1. .>639

FEATURES

source

STS
ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 639;
Best Local Similarity 92.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCAGTTCAGATA 25

Db 388 AAAAAAAAAAGTTCAGTTCAGTTA 412

RESULT 8
BV624515/c 777 bp DNA linear STS 15-Apr-2005
LOCUS 621586064FH6.T0 Clara Pan troglodytes troglodytes STS genomic,
DEFINITION sequence tagged site.
ACCESSION BV624515
VERSION BV624515.1 GI:62614693
KEYWORDS STS.
SOURCE Pan troglodytes troglodytes
ORGANISM Pan troglodytes troglodytes
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan (777)
1 (bases 1 to 777)
Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
Jaffe,D.B.
Initial Sequence of the Chimpanzee Genome and Comparison with the
Human Genome
Unpublished (2005)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Michael C. Zody
Broad Institute of MIT and Harvard
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172580933
Fax: 6172580903
Email: mczody@broad.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 777
Protocol:
23,021,928 chimpanzee whole genome shotgun reads were aligned to
the Human genome NCBI
Build 34 (Ng16,July 2003). Chimp WGS reads were from 9 donors,
including Clint (Pan
troglodytes verus), 3 other Pan troglodytes verus chimps
(Donald,Karlen,Yvonne), 3 Pan
troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
of unknown origin
(Don,Unknown Chimp). Common names: Pan troglodytes verus is the
western chimp and Pan
troglodytes troglodytes is the central chimp. To be included in
chimpanzee SNP discovery, a
read must be at least 500bp in length, at least 50% of its base
calls must have phred
score >= 20, at least 30% of its base calls must satisfy
SNOS(30,25) (single strand NGS, the
base in question has phred score >= 30, the surrounding 10 bases in
the read have phred
score >= 25), and the read must have at least 200 bp SNOS(30,25)
bases. Reads not uniquely
placed in the genome and read pairs whose two ends were not
consistently placed were
discarded. After above filtering, NGS(30,25) standard was applied
to all pairs of
overlapping reads to call NGS bases and SNPs. Alignments (between
two reads) with less
than 100 NGS bases or with SNP rate > 0.01 were discarded. To
exclude alignment between two
copies of a single read, comparisons between two reads that share
95% of their genome
alignments (>=95% bases of read A and >=95% bases of read B were
placed at the same locus
of human genome) were discarded.
Location/Qualifiers
1. 777
/organism="Pan troglodytes troglodytes"
/mol_type="genomic DNA"
/sub_species="troglodytes"
/db_xref="taxon:37011"

FEATURES

source

STS /clone.lib="Clara"
 <1..5777

Query Match 87.2%; Score 21.8; DB 10; length 777;
 Best Local Similarity 92.0%; Pred. No. 1.5e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 489 AAAAAAAAAAGTCCATACAGATA 465

RESULT 9
 AC141798/c
 LOCUS
 DEFINITION Apis mellifera clone CH224-61C22, *** SEQUENCING IN PROGRESS ***
 AC141798
 38 unordered pieces.
 AC141798.1 GI:29123982
 HTG: HTGS_PHASE1.
 KEYWORDS Apis mellifera (honey bee)
 SOURCE Apis mellifera
 ORGANISM Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preyigota;
 Neoptera; Endopterygota; Hymenoptera; Apoidea; Apoidea;
 Apidae; Apis.
 1 (bases 1 to 66834)
 Muny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alabrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Bivela,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,T., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carroo,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,X., Dinh,H.H.,
 Douhaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
 Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C.,
 Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichner,O., Lieu,C., Liu,U., Liu,W., Louised,H.,
 Lofado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okunolu,G.,
 Oreguine,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peter,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,
 Rivas,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sison,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Swatek,A., Tabor,P., Tamerias,A., Tamerias,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,F., Telford,B., Thomas,N., Thomas,S.,
 Umanu,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wlezyk,R., Woodson,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 66834)
 AUTHOR WORKLEY,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAR-2003) Human Genome Sequencing Center, Department

COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: AMCD
 Center clone name: CH224-61C22
 ----- Summary Statistics
 Sequencing vector: plasmid;
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 52310 bases at least Q40
 Consensus quality: 56858 bases at least Q30
 Consensus quality: 59773 bases at least Q20
 Estimated insert size: 48356; sum-of-coverage
 Quality coverage: 1x in Q20 bases; sum-of-coverage

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 38 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1092: contig of 1092 bp in length
 * 1093 1192: gap of unknown length
 * 1193 2380: contig of 1188 bp in length
 * 2381 2480: gap of unknown length
 * 2481 3973: contig of 1493 bp in length
 * 3974 4073: gap of unknown length
 * 4074 5236: contig of 1163 bp in length
 * 5237 5336: gap of unknown length
 * 5337 6574: contig of 1238 bp in length
 * 6575 7782: gap of unknown length
 * 7783 7882: contig of 1108 bp in length
 * 7883 9063: gap of unknown length
 * 9064 9163: contig of 1181 bp in length
 * 9164 10488: gap of unknown length
 * 10489 10588: contig of 1325 bp in length
 * 10589 11685: gap of unknown length
 * 11686 11785: contig of 1097 bp in length
 * 11786 13260: gap of unknown length
 * 13261 13360: contig of 1475 bp in length
 * 13361 14675: gap of unknown length
 * 14676 14775: contig of 1315 bp in length
 * 14776 16380: gap of unknown length
 * 16381 16480: contig of 1605 bp in length
 * 16481 17483: gap of unknown length
 * 17484 17583: contig of 1003 bp in length
 * 17584 18733: gap of unknown length
 * 18734 18833: contig of 1150 bp in length
 * 18834 20058: gap of unknown length
 * 20059 20158: contig of 1225 bp in length
 * 20159 21819: gap of unknown length
 * 21820 21919: contig of 1661 bp in length
 * 21920 23131: gap of unknown length
 * 23132 23231: contig of 1212 bp in length
 * 23232 24595: gap of unknown length
 * 24596 24959: contig of 1364 bp in length
 * 24960 25814: gap of unknown length
 * 25815 25914: contig of 1119 bp in length
 * 25915 27238: gap of unknown length
 * 27239 29299: contig of 1324 bp in length
 * 29300 29399: gap of unknown length
 * 29400 30778: contig of 1961 bp in length
 * 30779 30778: gap of unknown length
 * 30779 30778: contig of 1379 bp in length

RESULT 11
AC005152
LOCUS AC005152 126150 bp DNA linear PRI 27-JUN-1998
DEFINITION Homo sapiens chromosome 17, clone hRPC.1037_O_7, complete sequence.
ACCESSION AC005152
VERSION AC005152.1 GI:3264565
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 126150)
AUTHORS Birren, B., Fasmann, K., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 17, clone hRPC.1037_O_7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 126150)
AUTHORS Birren, B., Fasmann, K., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Baker, J., Baldwin, J., Barna, N., Beckert, R., Berni, J., Boatn, C.,
Boutwell, C., Brown, A., Casale, A., Cerny, J., Cooke, P., Depayre, E.,
Devon, K., Dewar, K., Donelan, L., Etemadi, S., Ferreira, P.,
FitzHugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S.,
Gensheimer, S., Geraghty, K., Gilmarin, T., Grant, G., Hagos, B.,
Harris, K., Horton, L., Howland, J. C., Hui, L., Jacotot, L., Kann, L.,
Meldrum, D., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Nechman, A., Nahl, R., Naylor, J., O'Connor, T., Pavlin, B.,
Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R.,
Stange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C.,
Strickland, C., Subramanian, A., Tornelli-Miller, I., Vassiliev, H.,
Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W. J., Zhao, J. and
Zody, M.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1998) Whitehead Institute/MIT Center for Genome
REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 126150)
Birren, B., Fasmann, K., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Baker, J., Baldwin, J., Barna, N., Beckert, R., Berni, J., Boatn, C.,
Boutwell, C., Brown, A., Casale, A., Cerny, J., Cooke, P., Depayre, E.,
Devon, K., Dewar, K., Donelan, L., Etemadi, S., Ferreira, P.,
FitzHugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S.,
Gensheimer, S., Geraghty, K., Gilmarin, T., Grant, G., Hagos, B.,
Harris, K., Horton, L., Howland, J. C., Hui, L., Jacotot, L., Kann, L.,
Meldrum, D., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Nechman, A., Nahl, R., Naylor, J., O'Connor, T., Pavlin, B.,
Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R.,
Stange-Thomann, N., Stilwell, J., Stojanovic, N., Stone, C.,
Strickland, C., Subramanian, A., Tornelli-Miller, I., Vassiliev, H.,
Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W. J., Zhao, J. and
Zody, M.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-1998) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 27, 1998 this sequence version replaced g1:355323.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source
location/Qualifiers
1..126150
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hRPC.1037_O_7"
/clone_lib="Peter de Jong/ human PAC library"
repeat_region
complement(1..492)
/rpt_family="L1PA13"
537..889
/rpt_family="THEIC"

repeat_region complement(1246..1579)
/rpt_family="L1MB3"
2011..2042
/rpt_family="POLY_A"
repeat_region complement(3417..3739)
/rpt_family="MER58B"
4080..4378
/rpt_family="ALUSx"
repeat_region complement(4505..5243)
/rpt_family="MER7B"
repeat_region complement(5251..5549)
/rpt_family="ALUSx"
repeat_region complement(5552..5863)
/rpt_family="MER7B"
repeat_region complement(6007..6034)
/rpt_family="(TAGA)n"
repeat_region complement(6037..6336)
/rpt_family="ALUY"
repeat_region complement(7279..7301)
/rpt_family="POLY_A"
repeat_region complement(7685..7810)
/rpt_family="FLAM_C"
7830..7929
/rpt_family="MIR"
repeat_region complement(9567..9767)
/rpt_family="L1MA2"
repeat_region complement(9764..9859)
/rpt_family="L1MA2"
10434..10559
/rpt_family="(TA)n"
repeat_region complement(10561..10866)
/rpt_family="ALUSg"
repeat_region complement(12551..13125)
/rpt_family="L1P4"
repeat_region complement(13253..14233)
/rpt_family="L1P4"
repeat_region complement(14246..14652)
/rpt_family="MSTB"
15011..15295
/rpt_family="ALUY"
repeat_region complement(15747..16046)
/rpt_family="ALUSx"
repeat_region complement(16090..16222)
/rpt_family="L12"
repeat_region complement(16394..16862)
/rpt_family="L12"
repeat_region complement(16868..16944)
/rpt_family="L12"
repeat_region complement(16905..16957)
/rpt_family="MIR"
repeat_region complement(17097..17305)
/rpt_family="L12"
repeat_region complement(17436..17631)
/rpt_family="L12"
repeat_region complement(21798..21837)
/rpt_family="(GA)n"
repeat_region complement(21950..22116)
/rpt_family="MIR"
23609..23910
/rpt_family="ALUY"
repeat_region complement(25116..25174)
/rpt_family="MER58B"
repeat_region complement(25147..25292)
/rpt_family="MER58B"
25437..25578
/rpt_family="L1MC/D"
repeat_region 25802..25995
/rpt_family="MIR"
repeat_region 26789..27094
/rpt_family="ALUSg"
27180..27203
/note="Single-stranded coverage."
repeat_region complement(27183..27468)

```

repeat_region      /rpt_family="AluSx"
                    complement(27637..27821)
repeat_region      /rpt_family="MIR"
                    complement(28663..28944)
repeat_region      /rpt_family="MLTID"
                    29165..29219
repeat_region      /rpt_family="MIR"
                    complement(29403..29701)
repeat_region      /rpt_family="AluY"
                    complement(29751..29791)
repeat_region      /rpt_family="AT_rich"
                    30826..30924
repeat_region      /rpt_family="POLY_A"
                    complement(32988..33141)
repeat_region      /rpt_family="L2"
                    33145..33230
repeat_region      /rpt_family="MER46"
                    33392..33467
repeat_region      /rpt_family="MER46"
                    33808..33908
repeat_region      /rpt_family="MLTIC"
                    33909..34268
repeat_region      /rpt_family="MLTIA1"
                    34269..34469
repeat_region      /rpt_family="MLTIC"
                    complement(34485..34783)
repeat_region      /rpt_family="AluSx"
                    34784..34920
repeat_region      /rpt_family="MLTIC"
                    complement(34975..35351)
repeat_region      /rpt_family="THEIB"
                    35718..37482
repeat_region      /rpt_family="L2"
                    complement(37515..37785)
repeat_region      /rpt_family="L2"
                    complement(37789..38175)
repeat_region      /rpt_family="MLTIB"
                    complement(38263..38561)
repeat_region      /rpt_family="AluSx"
                    complement(38563..38722)
repeat_region      /rpt_family="L2"
                    38734..38794
repeat_region      /rpt_family="MERS8B"
                    complement(38795..39051)
repeat_region      /rpt_family="THEIB"
                    complement(39348..39636)
repeat_region      /rpt_family="AluSx"
                    40932..41118
repeat_region      /rpt_family="MIR"
                    41241..41626
repeat_region      /rpt_family="L2"
                    complement(41649..41855)
repeat_region      /rpt_family="AluSx"
                    complement(43571..43598)
repeat_region      /rpt_family="AT_rich"
                    complement(44143..44245)
repeat_region      /rpt_family="AT_rich"
                    complement(44269..44514)
Query Match      87.2% Score 21.8; DB 8; Length 126150;
Best Local Similarity 92.0%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1 AAAAAAAAAAGTCCATTGAGATA 25
Db 76250 AAAAAAAAAAGTCCATTGAGATA 76274

```

```

RESULT 12 AC084783 166982 bp DNA linear PRI 13--JAN-2001
LOCUS AC084783
DEFINITION Homo sapiens chromosome 15 clone RP11-SN19 map 15q21.3, complete
sequence.

```

```

ACCESSION AC084783 AC015659
VERSION AC084783.2 GI:112203289
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominoidea; Homo.
REFERENCE 1 (bases 1 to 166982)
AUTHORS Dors,M., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
          Nesbitt,R., Traicoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
          Nesbitt,R., Traicoff,R. and Hood,L.
          Sequencing of human chromosome 15 D15S146-D15S117 region
          Unpublished
          2 (bases 1 to 166982)
REFERENCE Dors,M., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
          Nesbitt,R., Traicoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
          Nesbitt,R., Traicoff,R. and Hood,L.
          Direct Submission
          Submitted (16-NOV-2000) Multimegabase Sequencing Center, Institute
          for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
          98105, USA
          3 (bases 1 to 166982)
REFERENCE Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
          Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
          Nesbitt,R., Traicoff,R. and Hood,L.
          Direct Submission
          Submitted (13-JAN-2001) Multimegabase Sequencing Center, Institute
          for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
          98105, USA
          On Jan 13, 2001 this sequence version replaced gi:1181781.
COMMENT ----- Genome Center
          Center: Multimegabase Sequencing Center
          Center code: UWMSC
          Web site: http://chroma.mbt.washington.edu/msg_www
          Contact: leetownes@systembiology.org
          Drafting center: WIBR
          ----- Summary Statistics
          Sequencing vector: pUC18; L08752
          Chemistry: Dye-terminator Big Dye; 90% of reads
          Chemistry: Dye-primer Big Dye; 10% of reads
          Assembly program: Phrap; version 0.990399
          -----
          This clone was finished using overlapping sequence from AC036163
          [drafting center=WIBR] and AC068726 [drafting center = UWMSC]
          -----
FEATURES
source
Location/Qualifiers
1..166982
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.3"
/clone="RP11-SN19"
/clone="RP11-SN19"
/note="This clone overlaps RP11-241D5 AC036163 and
RP11-844G18 AC068726. Data from these BACs were combined
and the consensus sequence was derived from SN19 to the
extent possible."
1..54304
/note="overlap with RP11-241D5, AC036163"
47770..47790
/note="low quality data"
77830..78230
/note="low quality data"
82770..82790
/note="low quality data"
104080..104130
/note="low quality data"
159310..166982
/note="overlap with RP11-844G16, AC068726"
misc_feature
unsure
unsure
unsure
misc_feature
ORIGIN

```


TITLE
JOURNAL
COMMENT

Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teefaye,S., Theodore,J., Titzell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wymann,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 16, 2000 this sequence version replaced gi:7582614.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6526

Center clone name: 667_G_24

* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1071: contig of 1071 bp in length
* 1072 1171: gap of 100 bp
* 1172 2443: contig of 1272 bp in length
* 2444 2543: gap of 100 bp
* 2544 4110: contig of 1567 bp in length
* 4111 4210: gap of 100 bp
* 4211 5676: contig of 1466 bp in length
* 5677 6954: gap of 100 bp
* 6955 7054: gap of 100 bp
* 7055 8467: contig of 1413 bp in length
* 8468 8567: gap of 100 bp
* 8568 10046: contig of 1479 bp in length
* 10047 10146: gap of 100 bp
* 10147 11965: contig of 1819 bp in length
* 11966 12065: gap of 100 bp
* 12066 13754: contig of 1689 bp in length
* 13755 13854: gap of 100 bp
* 13855 15323: contig of 1369 bp in length
* 15324 16734: gap of 100 bp
* 16735 16834: gap of 100 bp
* 16835 18271: contig of 1437 bp in length
* 18272 18371: gap of 100 bp
* 18372 19546: contig of 1175 bp in length
* 19547 19646: gap of 100 bp
* 19647 20910: contig of 1264 bp in length
* 20911 21010: gap of 100 bp
* 21011 22952: contig of 1942 bp in length
* 22953 23052: gap of 100 bp
* 23053 24765: contig of 1713 bp in length
* 24766 24865: gap of 100 bp
* 24866 26182: contig of 1317 bp in length
* 26183 26282: gap of 100 bp
* 26283 27734: contig of 1452 bp in length
* 27735 27834: gap of 100 bp
* 27835 29209: contig of 1375 bp in length
* 29210 31590: contig of 2281 bp in length
* 31591 31690: gap of 100 bp
* 31691 33673: contig of 1983 bp in length
* 33674 33773: gap of 100 bp
* 33774 35628: contig of 1855 bp in length
* 35629 35728: gap of 100 bp
* 35729 37866: contig of 2138 bp in length

37867 37966: gap of 100 bp
* 37967 40549: contig of 2583 bp in length
* 40550 40649: gap of 100 bp
* 40650 42575: contig of 1926 bp in length
* 42576 42675: gap of 100 bp
* 42676 45614: contig of 2939 bp in length
* 45615 45714: gap of 100 bp
* 45715 48125: contig of 2411 bp in length
* 48126 48226: gap of 100 bp
* 48227 50634: contig of 2408 bp in length
* 50634 50733: gap of 100 bp
* 50734 53919: contig of 3186 bp in length
* 53920 54019: gap of 100 bp
* 54020 56516: contig of 2497 bp in length
* 56517 56616: gap of 100 bp
* 56617 59036: contig of 2420 bp in length
* 59037 59136: gap of 100 bp
* 59137 62691: contig of 3555 bp in length
* 62692 62791: gap of 100 bp
* 62792 65860: contig of 3069 bp in length
* 65861 65960: gap of 100 bp
* 65961 68910: contig of 2950 bp in length
* 68911 69010: gap of 100 bp
* 69011 72317: contig of 3307 bp in length
* 72318 72417: gap of 100 bp
* 72418 75535: contig of 3118 bp in length
* 75536 75635: gap of 100 bp
* 75636 78272: contig of 2637 bp in length
* 78273 78372: gap of 100 bp
* 78373 82077: contig of 3705 bp in length
* 82078 82177: gap of 100 bp
* 82178 86412: contig of 4235 bp in length
* 86413 86512: gap of 100 bp
* 86513 91421: contig of 4909 bp in length
* 91422 91521: gap of 100 bp
* 91522 96898: contig of 5377 bp in length
* 96899 96998: gap of 100 bp
* 96999 102230: contig of 5232 bp in length
* 102231 102330: gap of 100 bp
* 102331 106721: contig of 4391 bp in length
* 106722 106821: gap of 100 bp
* 106822 111047: contig of 4226 bp in length
* 111048 111147: gap of 100 bp
* 111148 116783: contig of 5636 bp in length
* 116784 116883: gap of 100 bp
* 116884 121808: contig of 4925 bp in length
* 121809 121908: gap of 100 bp
* 121909 127934: contig of 6026 bp in length
* 127935 128034: gap of 100 bp
* 128035 134776: contig of 6742 bp in length
* 134777 134876: gap of 100 bp
* 134877 142048: contig of 7172 bp in length
* 142049 142148: gap of 100 bp
* 142149 148878: contig of 6730 bp in length
* 148879 148978: gap of 100 bp
* 148979 157074: contig of 8096 bp in length
* 157075 157174: gap of 100 bp
* 157175 165913: contig of 8739 bp in length
* 165914 166013: gap of 100 bp
* 166014 177085: contig of 11072 bp in length
* 177086 177185: gap of 100 bp
* 177186 189278: contig of 12093 bp in length.

FEATURES

source

1.189278
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-667G24"
/clone_11b="RPC1-11 Human Male BAC"
1..1071
/note="assembly_fragment"

misc_feature


```

/note="<30 qual single clone coverage"
218..326
/note="<30 qual single clone coverage"
311
/note="probably C; possibly CC"
/rpt_family="B4A"
complement(1113..1382)
/note="single clone coverage"
complement(1121..1230)
/note="<30 qual single clone coverage"
complement(1128..1132)
/note="probably TTTT"
1246..1304
/rpt_family="(TC)n"
1477..1776
/rpt_family="MTE"
complement(1293..3178)
/rpt_family="ID_B1"
3068..3072
/note="<30 qual SNGL region"
3706..3739
/rpt_family="(TTTGG)n"
complement(3740..3888)
/rpt_family="B1_MM"
4098..4122
/rpt_family="(CA)n"
4150..4209
/rpt_family="AT_rich"
complement(4525..4668)
/rpt_family="ID_B1"
complement(4756..4934)
/rpt_family="MTD"
5385..5591
/rpt_family="B1_MM"
complement(5888..6240)
/rpt_family="MTE"
6433..6570
/rpt_family="B3A"
6589..6651
/rpt_family="(CA)n"
6652..6686
/rpt_family="AT_rich"
6823..7050
/rpt_family="B1_MM"
7054..7244
/rpt_family="ORR1D"
7748..7853
/rpt_family="ORR1C"
7993..8117
/rpt_family="B1_MM"
8118..8145
/rpt_family="B4A"
8851..9039
/rpt_family="B2_MM2"
9066..9220
/rpt_family="MTC"
9339..9428
/rpt_family="(TTTC)n"
9885..9923
/rpt_family="(TTGG)n"
9925..9970
/rpt_family="(TTTGG)n"
9986..10288
/rpt_family="MTC"
10805..10893
/rpt_family="B3A"
10917..10969
/rpt_family="(CA)n"
10971..10992
/rpt_family="AT_rich"
complement(11474..11884)
/rpt_family="RMR15"

```

```

repeat_region complement(12604..13182)
/rpt_family="L1"
repeat_region 13187..13345
/rpt_family="(TTC)n"
repeat_region 13353..13411
/rpt_family="(TTTGA)n"
repeat_region complement(13412..14146)
/rpt_family="L1"
repeat_region 14147..14183
/rpt_family="(TTG)n"
repeat_region complement(14184..14666)
/rpt_family="L1"
repeat_region 14667..14705
/rpt_family="(TTTA)n"

Query Match 87.2% Score 21.8; DB 9; Length 199808;
Best Local Similarity 92.0%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 173082 AAAAAAAAAAGTTCATTCAGATA 173058

RESULT 16
AC164224
LOCUS
DEFINITION
Bos taurus clone CH240-127C15, *** SEQUENCING IN PROGRESS ***, 41
unordered pieces.
AC164224 220832 bp DNA linear HTG 01-JUL-2005
AC164224.2 GI:68300957
VERSION
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Bos taurus (cow)
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 220832)
Muzny,D,Marle, Metzker,M,Lee, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Albrechts,S, Amin,A, Angiano,D,
Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Buttrill,K, Calderon,S,
Cardenas,V, Carter,K, Cavazos,I, Cessari,H, Canter,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyte,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denison,S, Detamo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K,
Egan,A, Escoto,M, Eugene,C, Evans,C,A, Falle,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Geisregeorgie,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunaratne,P, Haaland,M, Hamill,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogue,M,
Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kovis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorenshaw,W, Louisedge,H, Lozada,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindaratne,M, Mahmood,M, Malloy,K, Mangum,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
McInerney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minde,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
Nwokeneme,O, Okwou,G, Olarpunsaogun,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,J, Pfannkuch,C,
Plopper,F, Poidexter,A, Popovic,D, Primus,E, Pu,L,
Punzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,

```


Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajic, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Wright, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstein, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 220832)
 Morley, K.C.
 Direct Submission
 Submitted (18-JUN-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 220832)
 Cow Genome Sequencing Consortium.
 Direct Submission
 Submitted (01-JUL-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 29, 2005 this sequence version replaced gi:67972531.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: RPN
 Center clone name: CH240-127C15
 ----- Summary Statistics
 Assembly program: Atlas 3.0
 Consensus quality: 197006 bases at least Q40
 Consensus quality: 200464 bases at least Q30
 Consensus quality: 203406 bases at least Q20
 Estimated insert size: 201090; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	3815: contig of 3815 bp in length
3816	4147: gap of 332 bp
4148	9632: contig of 5485 bp in length
9633	11072: gap of 1440 bp
11073	12374: contig of 1302 bp in length
12375	12474: gap of unknown length
12475	14360: contig of 1886 bp in length
14361	14410: gap of 50 bp
14411	17251: contig of 2841 bp in length
17252	17301: gap of 50 bp
17302	19953: contig of 2652 bp in length
19954	20003: gap of 50 bp
20004	25819: contig of 5816 bp in length
25820	25869: gap of 50 bp
25870	36818: contig of 10948 bp in length
36819	36867: gap of 50 bp
36868	43023: contig of 6155 bp in length
43024	43072: gap of 50 bp
43073	45717: contig of 2645 bp in length
45718	45767: gap of 50 bp
45768	47912: contig of 2145 bp in length
47913	48503: gap of 591 bp
48504	56785: contig of 8282 bp in length
56786	56835: gap of 50 bp
56836	60521: contig of 3686 bp in length
60522	60571: gap of 50 bp
60572	69004: contig of 8433 bp in length
69005	69054: gap of 50 bp
69055	74514: contig of 5460 bp in length
74515	74564: gap of 50 bp
74565	80002: contig of 5438 bp in length
80003	80052: gap of 50 bp
80053	82765: contig of 2713 bp in length
82766	82815: gap of 50 bp
82816	86676: contig of 3861 bp in length
86677	86728: gap of 50 bp
86729	94359: contig of 7633 bp in length
94360	95045: gap of 686 bp
95046	98253: contig of 3208 bp in length
98254	98303: gap of 50 bp
98304	122673: contig of 24370 bp in length
122674	122723: gap of 50 bp
122724	132808: contig of 10085 bp in length
132809	132858: gap of 50 bp
132859	135630: contig of 2772 bp in length
135631	135680: gap of 50 bp
135681	141950: contig of 6270 bp in length
141951	142000: gap of 50 bp
142001	146191: contig of 4191 bp in length
146192	146241: gap of 50 bp
146242	157227: contig of 10966 bp in length
157228	157277: gap of 50 bp
157278	159092: contig of 1815 bp in length
159093	167105: gap of 8013 bp
167106	169687: contig of 2582 bp in length
169688	169787: gap of unknown length
169788	171270: contig of 1483 bp in length
171271	171320: gap of 50 bp
171321	179856: contig of 8536 bp in length
179857	179956: gap of unknown length
179957	181796: contig of 1840 bp in length
181797	181846: gap of 50 bp
181847	191335: contig of 9489 bp in length
191336	191435: gap of unknown length
191436	192671: contig of 1236 bp in length
192672	192771: gap of unknown length
192772	193856: contig of 1085 bp in length
193857	193956: gap of unknown length
193957	195144: contig of 1188 bp in length
195145	195244: gap of unknown length
195245	196751: contig of 1507 bp in length
196752	196851: gap of unknown length
196852	198048: contig of 1197 bp in length
198049	198148: gap of unknown length
198149	199169: contig of 1021 bp in length
199170	199269: gap of unknown length
199270	202286: contig of 3017 bp in length
202287	202386: gap of unknown length
202387	206603: contig of 4217 bp in length

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

* 206604 206703: gap of unknown length
 * 206704 220832: contig of 14129 bp in length.
 Location/Qualifiers

1..220832
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"

Query Match 87.2%; Score 21.8; DB 14; Length 220832;
 Best Local Similarity 92.0%; Pred. No. 2.4e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTGAGA 25
 Db 85148 AAAAAAAAAATTCCATTGAGTA 85172

RESULT 17
 AB128049_15/c
 WPCOMMENT

Sequence split into 33 fragments, LOCUS AB128049 Accession AB128049

Fragment Name	Begin	End
AB128049_00	1	110000
AB128049_01	100001	210000
AB128049_02	200001	310000
AB128049_03	300001	410000
AB128049_04	400001	510000
AB128049_05	500001	610000
AB128049_06	600001	710000
AB128049_07	700001	810000
AB128049_08	800001	910000
AB128049_09	900001	1010000
AB128049_10	1000001	1110000
AB128049_11	1100001	1210000
AB128049_12	1200001	1310000
AB128049_13	1300001	1410000
AB128049_14	1400001	1510000
AB128049_15	1500001	1610000
AB128049_16	1600001	1710000
AB128049_17	1700001	1810000
AB128049_18	1800001	1910000
AB128049_19	1900001	2010000
AB128049_20	2000001	2110000
AB128049_21	2100001	2210000
AB128049_22	2200001	2310000
AB128049_23	2300001	2410000
AB128049_24	2400001	2510000
AB128049_25	2500001	2610000
AB128049_26	2600001	2710000
AB128049_27	2700001	2810000
AB128049_28	2800001	2910000
AB128049_29	2900001	3010000
AB128049_30	3000001	3110000
AB128049_31	3100001	3210000
AB128049_32	3200001	3284914

Continuation (16 of 33) of AB128049 from base 1500001 (AB128049 Macaca mulatta genes, MT)

Query Match 85.6%; Score 21.4; DB 8; Length 110000;
 Best Local Similarity 95.7%; Pred. No. 4.1e+02;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTGAGA 23
 Db 98561 AAAAAAAAAAGTTCATTGAGA 98539

RESULT 18
 AC148704 147539 bp DNA linear PRI 12-APR-2004
 LOCUS AC148704
 DEFINITION Macaca mulatta Major Histocompatibility Complex BAC MMU337E21,
 complete sequence.
 AC148704
 VERSION AC148704.1 GI:46358486

KEYWORDS
 SOURCE
 ORGANISM

HTG.
 Macaca mulatta (rhesus monkey)
 Macaca mulatta

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE
 AUTHORS
 TITLE
 Daza-Vamanta, R., Guthrie, B., Glusman, G., Rowan, L., Vu, Q.,
 Williams, L.M., Kurihara, B.K. and Geraghty, D.E.
 Immunological Divergence Between Rhesus Monkeys (Macaca mulatta)
 and Human as Revealed by the Complete Sequence of a Macaque Major
 Histocompatibility Complex (MHC)

JOURNAL
 REMARK
 The BAC is one of 59 projects that collectively comprise a rhesus
 macaque major histocompatibility complex genomic sequence.
 Reference to the annotated contiguous sequence

REFERENCE

2 (bases 1 to 147539)
 Daza-Vamanta, R., Guthrie, B., Glusman, G., Rowan, L., Vu, Q.,
 Williams, L.M., Kurihara, B.K. and Geraghty, D.E.
 Direct Submission
 Submitted (12-APR-2004) Clinical Research, Fred Hutchinson Cancer
 Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA
 Location/Qualifiers

FEATURES

1..147539
 /organism="Macaca mulatta"
 /mol_type="genomic DNA"
 /db_xref="taxon:9544"
 /clone="BAC MMU337E21"
 /clone_1lb="The CHORI-250 Rhesus macaque BAC library
 constructed at the Children's Hospital Oakland Research
 Institute, BACPAC Resources"

ORIGIN

Query Match 85.6%; Score 21.4; DB 8; Length 147539;
 Best Local Similarity 95.7%; Pred. No. 3.7e+02;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTGAGA 23
 Db 28962 AAAAAAAAAAGTTCATTGAGA 28984

RESULT 19

AC148710 158818 bp DNA linear PRI 12-APR-2004
 LOCUS AC148710
 DEFINITION Macaca mulatta Major Histocompatibility Complex BAC MMU39F22,
 complete sequence.
 AC148710
 VERSION AC148710.1 GI:46358492

KEYWORDS

HTG.
 Macaca mulatta (rhesus monkey)
 Macaca mulatta

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE
 AUTHORS
 TITLE
 Daza-Vamanta, R., Guthrie, B., Glusman, G., Rowan, L., Vu, Q.,
 Williams, L.M., Kurihara, B.K. and Geraghty, D.E.
 Immunological Divergence Between Rhesus Monkeys (Macaca mulatta)
 and Human as Revealed by the Complete Sequence of a Macaque Major
 Histocompatibility Complex (MHC)

JOURNAL

Unpublished
 The BAC is one of 59 projects that collectively comprise a rhesus
 macaque major histocompatibility complex genomic sequence.
 Reference to the annotated contiguous sequence

REFERENCE

2 (bases 1 to 158818)
 Daza-Vamanta, R., Guthrie, B., Glusman, G., Rowan, L., Vu, Q.,
 Williams, L.M., Kurihara, B.K. and Geraghty, D.E.
 Direct Submission
 Submitted (12-APR-2004) Clinical Research, Fred Hutchinson Cancer
 Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA
 Location/Qualifiers

FEATURES

1..158818

mRNA 62955. .63878
/gene="MGCI0955"
CDS 63167. .63262
/gene="MGCI0955"
/note="Homo sapiens hypothetical protein MGCI0955 (MGCI0955), mRNA.; H_NH0287D01.2
This gene was based on gi(32261298) "
/codon_start=1
/product="unknown"
/protein_id="AA03058.1"
/db_xref="GI:62702131"
/translation="MSKTEPAAENDKASVAGQITQTGP"
/gene="tmp_locus_40"
complement(63898. .76391)
/gene="tmp_locus_40"
complement(join(63898. .64141,70755. .70843,73672. .73786,76304. .76357))
/gene="tmp_locus_40"
complement(join(64076. .64141,70755. .70843,73672. .73786,76304. .76357))
/gene="Homo sapiens cDNA clone MGC:48612 IMAGE:4829356, complete cds.; H_NH0287D01.3
This gene was based on gi(27469510) "
/codon_start=1
/product="unknown"
/protein_id="AA03059.1"
/db_xref="GI:62702132"
/translation="MAAMSPAAAPLRLGIRGLPHHRMPATOTEGELRYTOLIKKEFPRATIKVTDSGCCGAMYEIKISEBEKERTVQCHOMVQALKEIKEMGLRITTSVPRK"
misc_feature 76195. .76613
/note="CpG island (%GC=79.5, c/e=0.94, #CpGs=58) "
complement(83343. .107340)
/gene="C2orf6"
complement(join(83343. .85128,87574. .87737,93701. .93834,95497. .95590,101069. .101235,107144. .107340))
/gene="C2orf6"
complement(join(85051. .85128,87574. .87737,93701. .93834,95497. .95590,101069. .101235,107144. .107157))
/gene="C2orf6"
/note="Homo sapiens chromosome 2 open reading frame 6 (C2orf6), mRNA.; H_NH0287D01.4
This gene was based on gi(8922670) "
/codon_start=1
/product="unknown"
/protein_id="AA03060.1"
/db_xref="GI:62702133"
/translation="MSFLFSRSRSTKPKKNIPESGHOYELKHAATIGSGNLRQAVMPEGDLNEMIAVNTVDPEFNQINMLYGTTERCTEASCPVMSAGRYEHMADGTAIKKPIKCSAPRYIDYLMTWQDQJLDELTLPSPKGVPEPKNFMVSVAITIKRLFRVVAHIYHQHDSVWQLQOEFAHLNLSFGHIFVQEFNLIIDRELAIPQELIEKLGSKDR"
misc_feature 102155. .102427
/note="CpG island (%GC=61.2, c/e=0.95, #CpGs=21) "
106709. .107520
/note="CpG_island (%GC=66.9, c/e=0.98, #CpGs=84) "
126801. .127779
/note="CpG_island (%GC=68.8, c/e=0.85, #CpGs=109) "
127068. .143768
/gene="MTHFD2"
join(127068. .127225,134188. .134372,136187. .136309,137052. .137204,138425. .138533,139682. .139774,140224. .140349,142562. .143768)
/gene="MTHFD2"
join(127143. .127225,134188. .134372,136187. .136309,137052. .137204,138425. .138533,139682. .139774,140224. .140349,142562. .142722)
/gene="MTHFD2"
/note="Homo sapiens methylene tetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase (MTHFD2), nuclear gene encoding mitochondrial protein, mRNA.; H_NH0287D01.5
This gene was based on gi(13699869) "
/codon_start=1

/product="unknown"
Query Match 85.6%; Score 21.4; DB 8; Length 183279;
Best Local Similarity 95.7%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Cy 1 AAAAAAAAAAGTTCATTCAGA 23
|||||
Db 87271 AAAAAAAAAAGTTCATTCAGA 87293
|||||
RESULT 21
AC148705 186491 bp DNA linear PRI 12-APR-2004
LOCUS Macaca mulatta Major Histocompatibility Complex MMU348N13, complete
DEFINITION sequence.
ACCESSION AC148705.1 GI:46358487
VERSION AC148705
KEYWORDS HTG.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecoidea; Macaca.
REFERENCE 1 (bases 1 to 186491)
Daza-Vamanta, R., Guthrie, B., Glusman, G., Rowan, L., Vu, Q.,
Williams, L.M., Kurihara, B.K. and Geraghty, D.E.
Immunological divergence between Rhesus Monkeys (Macaca mulatta)
and Human as revealed by the complete sequence of a Macaque Major
Histocompatibility Complex (MHC)
Unpublished
JOURNAL The BAC is one of 59 projects that collectively comprise a rhesus
REMARK macaque major histocompatibility complex genomic sequence.
REFERENCE 2 (bases 1 to 186491)
Daza-Vamanta, R., Guthrie, B., Glusman, G., Rowan, L., Vu, Q.,
Williams, L.M., Kurihara, B.K. and Geraghty, D.E.
Direct Submission
TITLE Submitted (12-APR-2004) Clinical Research, Fred Hutchinson Cancer
JOURNAL Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA
FEATURES
source location/Qualifiers
1. 186491
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/db_xref="taxon:9544"
/clone="BAC MMU348N13"
/clone_1lb="The CHORI-250 Rhesus macaque BAC library
constructed at the Children's Hospital Oakland Research
Institute, BACPAC Resources"
ORIGIN
Query Match 85.6%; Score 21.4; DB 8; Length 186491;
Best Local Similarity 95.7%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Cy 1 AAAAAAAAAAGTTCATTCAGA 23
|||||
Db 127493 AAAAAAAAAAGTTCATTCAGA 127515
|||||
RESULT 22
AC159797 209056 bp DNA linear HTG 01-JUL-2005
LOCUS Bos taurus clone CH240-76E11, *** SEQUENCING IN PROGRESS ***, 8
DEFINITION unordered pieces.
ACCESSION AC159797
VERSION AC159797.2 GI:68301208
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.

REFERENCE
AUTHORS

1 (bases 1 to 209056)

Munry, D. Marie, Metzger, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anulepcich, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bialwalo, K., Blat, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Diaper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gevorgyan, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Louise, H., Lora, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mamoud, M., Malloy, K., Mangun, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Miliostavljevic, A., Miner, G., Minja, E., Montemayor, J., Mollasavlis, A., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, G., Nguyen, N., Norris, S., Nwokedi, O., Okwum, G., Olarinmaga, A., Pal, S., Parks, K., Paeternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valias, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczek, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weise, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G., and Gibbs, R. A.

TITLE
JOURNAL

2 (bases 1 to 209056)

Worley, K. C.

Direct Submission

Submitted (09-APR-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

On Jun 29, 2005 this sequence version replaced gi:62460730.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: FEHT

Center clone name: CH240-76E11

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 205315 bases at least Q40

Consensus quality: 207032 bases at least Q30

Consensus quality: 207798 bases at least Q20

Estimated insert size: 209550; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 28385: contig of 28385 bp in length
* 28386 28435: gap of 50 bp
* 28436 89788: contig of 61532 bp in length
* 89789 89838: gap of 50 bp
* 89839 106859: contig of 17021 bp in length
* 106860 106909: gap of 50 bp
* 106910 124729: contig of 17820 bp in length
* 124730 124779: gap of 50 bp
* 124780 204917: contig of 80138 bp in length
* 204918 206057: gap of unknown length
* 206058 206157: contig of 1040 bp in length
* 206158 207209: gap of unknown length
* 207210 207309: contig of 1052 bp in length
* 207310 209056: contig of 1747 bp in length.

FEATURES
SOURCE

Location/Qualifiers

1. 209056

/organism="Bos taurus"

/mol_type="genomic DNA"

/db_xref="taxon:9913"

/clone="CH240-76E11"

28386..28435

/estimated_length=50

89789..89838

/estimated_length=50

106860..106909

/estimated_length=50

124730..124779

/estimated_length=50

204918..205017

/estimated_length=unknown

206058..206157

/estimated_length=unknown

207210..207309

/estimated_length=unknown

ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 209056;

Best Local Similarity 100.0%; Pred. No. 4.5e+02;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAAAAAAAAAGTTCATTCA 21

|||||

Db 168772 AAAAAAAAAAGTCCATTCA 168752

RESULT 23
 HS214XG7/c 307 bp DNA linear PRI 28-NOV-1994
 LOCUS H. sapiens (U1151316) DNA segment containing (CA) repeat; clone
 DEFINITION AF214xg7, single read.
 ACCESSION Z23630
 VERSION Z23630.1 GI:393830
 KEYWORDS CA repeat; dinucleotide repeat; GT repeat; microsatellite DNA;
 SOURCE microsatellite marker; repeat polymorphism.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.

REFERENCE
 1 Gypay,G., Morissette,J., Vignal,A., Dib,C., Fizames,C.,
 Millaesau,P., Marc,S., Bernardi,G., Lathrop,M. and Weissenbach,J.
 The 1993-94 Genethon human genetic linkage map
 Nat. Genet. 7 (2 Spec No), 246-339 (1994)

JOURNAL
 PUBMED 7545953
 2 (bases 1 to 307)
 Weissenbach,J.
 Direct Submission
 Submitted (12-JUL-1993) Genethon, B.P. 60, 91002 Evry Cedex France.
 E-mail: Jean.Weissenbach@genethon.fr
 Cloning vector is M13mp18ASB; full automatic.

COMMENT
 Location/Qualifiers
 1..307
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /cell_line="CEPH 134702"
 /clone_lib="genomic DNA"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 8; Length 307;
 Best Local Similarity 91.7%; Pred. No. 4.6e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTCAATTCAGAT 24
 |||||
 245 AAAAAAAAAAGTCCACTCAGAT 222
 |||||

RESULT 24
 BV003898 421 bp DNA linear STS 07-SEP-2002
 LOCUS S209P6140RH9.T0 C3H/HeJ Mus musculus STS genomic, sequence tagged
 DEFINITION site.
 ACCESSION BV003898
 VERSION BV003898.1 GI:22756975
 KEYWORDS STS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 421)
 Wade,C.
 Polymorphism Structure in the Mouse
 Unpublished (2002)

REFERENCE
 AUTHORS Kerstin Lindblad-Toh
 TITLE Whitehead Institute for Biomedical Research, Center for Genome
 Research
 320 Charles Street, Cambridge, MA 02141, USA
 Tel: 6172521477
 Fax: 6172580903

Email: kersli@genome.wi.mit.edu
 Primer A: None
 Primer B: None
 STS size: 421
 Protocol:
 WGS-discovery: Paired-end low-coverage whole genome shotgun reads
 were generated from 129S1/SVIMJ, C3H/HeJ, and BALB/cByJ. The WGS
 reads were placed uniquely on the MGSVC3 C57BL/6J assembly and SNP
 detection was carried out by SSAHA-SNP. 225,000 reads were
 annotated
 as STS and 81,000 SNPs were annotated with alleles from C57BL/6J
 and the strain from which the particular read came. The validation
 rate for these SNPs was estimated at approximately 98%.

FEATURES
 source
 1..421
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C3H/HeJ"
 /db_xref="taxon:10090"
 /map="2 22-370 128712564-128712909"
 /clone_lib="C3H/HeJ"
 <1..>421

ORIGIN
 STS
 Query Match 83.2%; Score 20.8; DB 10; Length 421;
 Best Local Similarity 91.7%; Pred. No. 4.1e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTCCATTCAATTCAGATA 25
 |||||
 128 AAAAAAAAAAGTCCACTCAGATA 151
 |||||

RESULT 25
 AB183841/c 4264 bp DNA linear BCT 26-JAN-2005
 LOCUS Streptococcus suis purH, seps16A, seps16B, seps16C, purH genes,
 DEFINITION partial and complete cds, strain:2726.
 ACCESSION AB183841
 VERSION AB183841.1 GI:58197481
 KEYWORDS Streptococcus suis
 SOURCE Streptococcus suis
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.

REFERENCE
 1 Sekizaki,T., Takamatsu,D., Oaki,M. and Shimoi,Y.
 Different foreign genes incidentally integrated into the same locus
 of the Streptococcus suis Genome
 J. Bacteriol. 187 (3), 872-883 (2005)

JOURNAL
 PUBMED 1559665
 2 (bases 1 to 4264)
 Sekizaki,T., Takamatsu,D., Oaki,M. and Shimoi,Y.
 Direct Submission
 Submitted (13-JUL-2004) Teutomu Sekizaki, National Institute of
 Animal Health, Molecular Bacteriology Section, Kannondai 3-1-5,
 Tsukuba, Ibaraki, 305-0856, Japan (E-mail:sekizaki@affrc.go.jp,
 Tel:81-29-838-7743, Fax:81-29-838-7907)

FEATURES
 source
 1..4264
 /organism="Streptococcus suis"
 /mol_type="genomic DNA"
 /strain="2726"
 /serotype="16"
 /db_xref="taxon:1307"
 1..537
 /gene="purH"
 <1..537
 /gene="purH"
 /BC number="6.3.2.6"
 /codon_start=1
 /evidence=not_experimental
 /translation_table=11

ORIGIN	Query Match	83.2%	Score 20.8	DB 1	Length 4264
Best Local Similarity	91.7%	Prod. No. 1.9e+03			
Matches 22	Conservative 0	Mismatches 2	Indels 0	Gaps 0	
QY	2	AAAAAAAAAGTTCAAATTCAGATA 25			
Db	3440	AAAACAAAGTTCATCGCAGATA 3417			
RESULT 26					
LOCUS	AE011418/c				
DEFINITION	AE011418	13356 bp	DNA	linear	BCT 28-Apr-2003
ACCESSION	AE011418				
VERSION	AE010300				
KEYWORDS	AE011418.1	GI:24196285			
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
PUBMED					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
FEATURES					
source					
gene					
CDS					

gene
/product="hief protein"
/protein_id="AAN49705.1"
/db_xref="GI:24196286"
/translation="MSNLTA RVLPCLIDKGRVYGVFNVLVDAGDPVSAALYEEN
LADPLCFIDITASDREELIHLYERLA EKIFIDPPTVGGIRTYDDPKALIEGADV
SINTAARONPKLITVSSSEITVGSOCITCAIDVKEKENDREVEFLHGRTITGREALD
AQEAERAGAGELITLSDMDRGTNRGPDHILKNSSSLGIPITLASGAGNPEHVEAI
LRGADAVLAASIFHFGESYIRRETKAMEEWGIVRLD"
complement(1111. .2574)
/gene="gata"
/locus_tag="LA2507"
complement(1111. .2574)
CDS
/locus_tag="LA2507"
/gene="gata"
/locus_tag="LA2507"
/EC_number="6.3.5.-"
/note="confirmed by proteomics"
/codon_start=1
/transl_table=1
/product="glutamyl-ERNAgl amidotransferase subunit A"
/protein_id="AAN49706.1"
/db_xref="GI:24196287"
/translation="MNRILKSYVELKDSLNSGKISSSTELVASACINRIRVDGAVKAF
LYDEKYLNSAESDKRKRGKPLSEPDGPVPAIKDNICIRDTITSCSKLIENTYS
PFHATVVEKLEKGFILPPTNMDSEFAMGSTENSAQTENPDLERIRGSGSGSA
AAVASWVPLAGSDTGSVQAPASLGLYGLKPTYGTSRYGLVAYASSLDQIGPS
RELQGCIDLVSIGKDVSDTSIHRPKFSADVPQDFKGLKGVIKMPEIQSEV
KSYKVLNOLKEKATLVLDLDFSKFGAIPYVYIIATAECNSLSPDGFIRFGSRKX
TGLKEDLPVDSRTSGPGEVDFRILITGTFSLAGYDAYGTAQAKARALLRKESEPF
SKVDCITIQPTSPPTAFKIGKTRDPIOMYKADITVTSVNLAGLPAMSVPMGADQGLP
IGLIQITPHFQEGKLFGLALSTLEGNTIQFPSSIK"
complement(2571. .2861)
/gene="gac"
/locus_tag="LA2508"
complement(2571. .2861)
CDS
/locus_tag="LA2508"
/gene="gac"
/locus_tag="LA2508"
/EC_number="6.3.5.-"
/codon_start=1
/transl_table=1
/product="glu-ERNAgl amidotransferase subunit C"
/protein_id="AAN49707.1"
/db_xref="GI:24196288"
/translation="MNINEDSLQIAELSRUKIRSEKEATLQDFNKILEYDQIKGL
DVSSIKDDEIYLHHEHNAIRPDLAKHLSREIESFAPSFGVGVFVVKVLET"
complement(2888. .4309)
/gene="rfb2"
/locus_tag="LA2509"
complement(2888. .4309)
CDS
/locus_tag="LA2509"
/gene="rfb2"
/locus_tag="LA2509"
/codon_start=1
/transl_table=1
/product="putative colanic biosynthesis UDP-glucose lipid
carrier transferase"
/protein_id="AAN49708.1"
/db_xref="GI:24196289"
/translation="MIMKERSQTFKLIFTLLDFSNALFSGILAFVFRFVPLDENGTD
RRVYDISEYILFLFLAFFOVYLAIDLVHPRGLSIDFLITIGSVFNLVILVA
VLFFRGDLGSRERSRYVLAFAMINITSSILHYIARIYIRILARKGYNIRSLVVG
VSETSKRPNDVAVIHGIYKILGIVQTKSAKPSKENVGKLEKISHVLEKRPDL
VVYTLPEBGDYLVDACDHGEGIDLIVGFQEFIKARGRVEEMGLPVISIRNIP
IRLYGNKVKRIPDLTSELILFLFSPPYLMALVLKLTSPGVFYOEIRVGLNKKF
KMIRKRTVVOAKQSEETTVQNDPRVTSVGRILRKLSLDETPOFENVLFGDSVVG
PRPRPHVEKFKNDKHVMRRHVAVKIGITGLAQVKGIRGDTSEDRIADIVYIEMW
SIMDLKILLTPKXGIMDKAY"
complement(4306. .7281)
/locus_tag="LA2510"
complement(4306. .7281)
CDS
/locus_tag="LA2510"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAN49709.1"

gene
/db_xref="GI:24196290"
/translation="MPLRVLILSLILIKIPVIAOSTDEQTLNLITGAGSDSSDDK
KALILAKRNLIGKSIDTLPDERIDLLSLIGTSGSLPTRRRKLRAALEPLPV
DPMVQIPOTKALPISIBNASEGELLQVNDKGVILRVLRVAKLKRSGSLAEKITV
DSRROSTIYAEGLIVYDGAIVYKDKLITDFREKVVYVTKTGFPAHITGKLLK
DDKRYALMGVFTICNAEKPHYFVRKNTLYITEDKYMAINVAIVYQGGTVLMPLY
NSLVGQWIAQAQKKNTOGLFLOTSTOWSTIPPLAPMOKYKTRADYEITGGAFFHS
MMQSPALNVLIDIAAYANHRVQTTAYEFRNFNGVGTAAVNOVDRGLYSADPNA
PLRNIGDVEPMWKGRIILNSKTNTEKEDTRNISIQYENYTDRLDEYEGNREYPA
TLQSLYARVRMGFTIRNTLEMLDYTENGDGINSIGMKRILLYILNADKSGYFP
TYDITPTTIRNSSEIGRIPYNSAVYMDVFNLTILRYGAPIRLNLKIPITDGSRC
DPMGDYKENVFRTOYITQSSGRLTINBSYLAFTPNAPFGAKQSANRNTAYTG
VTNAPTSLERYLARSEYELRTSSNIRFQIPPLFETTYRKTEIYKELQDPLTGK
RQHELELSLEVALENFEISIRTRIDLRSSDEKQPTDAERNYFVARSGYFDL
DGLRPKRVSLERKRSFYSGLFLANDYVHNPAKPLYNLSLTASVKAGGFTLPFRLI
RELLEGGTWHVYNSPILDGVRVYKANVDFTYLGLEALDSRVSGQPMRYTNOVNT
YDTPFYGNDDPTMSVATNLNDRSLQKDLIDGTVNGSGRONTALNINRPMGITKYL
HTMFRUGYSMDRSVPQGRITDGLVSEFYDSVFFSISITDPTIGQDSSSLTRYLR
FRRPFOAGSDTGISSEN"
complement(7305. .8138)
/locus_tag="LA2511"
complement(7305. .8138)
CDS
/locus_tag="LA2511"
/EC_number="2.7.1.66"
/note="putative undecaprenol kinase"
/codon_start=1
/transl_table=1
/product="Bactitracin resistance protein"
/protein_id="AAN49710.1"
/db_xref="GI:24196291"
/translation="MNPYLAFLRSITIEATFELPVSSTGHLPLFSFPFPGYENVE
DDLFDIFIQGAALISVFLYREKFSQIVSFPYILIKENYDSRQFHLQICIGAPFI
LIAGFLAKFLDITIKARPDLLSILSGMIGVGLIIVAEVYFQRPREKSGIGKDI
LIGIQOMALIPGMSREARTITARFLRGKRTQSAESEFLAVPVLAAIGYLIKRR
SIANGNTIYLMGFLVSRFLCTLVIRWFLRYIQKKSFSFGYVRIILGQVLYLTLL
I"
complement(8145. .8870)
/locus_tag="LA2512"
complement(8145. .8870)
CDS
/locus_tag="LA2512"
/note="confirmed by proteomics"
/codon_start=1
/transl_table=1
/product="unknown protein"
/protein_id="AAN49711.1"
/db_xref="GI:24196292"
/translation="MKNSITLIFVSLITAFASCGDNSEVIENTLGNKITYNSFEDY
NVALDMSRVQNTKEKULEFISKDISVEPOMRALNYQOKKRYTOYQDMATTITA
AEKDGFTKRDDIKKILKFEQMOQIVSQLYVHLVSKIKIISBEAMECQKLRSEPGI
SSLPIDRCILFARAKLKKDKSOEITLPVLERIKQVAKINDKFDLDAFLKKIKGBA
DKKSSSATGTAPKTEPXTTQ"
complement(8917. .12477)
/locus_tag="LA2513"
complement(8917. .12477)
CDS
/locus_tag="LA2513"
/codon_start=1
/transl_table=1
/product="transcription-repair coupling factor"
/protein_id="AAN49712.1"
/db_xref="GI:24196293"
/translation="WKDLNNIGDGVFSRBEBSLKKKNSITGKSKSPADSNNLTKQSN
LITLTKQASASEKKISVNTVTVGVSVYVTTGSHSILASSLFOKLNQITLIVSENNTA
EPLFRLASLPLSNMILYLGOEVLPEYMRPYSEMRERIKALAKLISGPPVLPIS
VSGFLKTLPIQMOGAAIYLKKGKEDLBSLLQLIDLGKRVQVCEFTQEPALKQ
ILDPSYSSTEPVRLDIFGEIEISIRFPEDDSQSMMDLQOAVLPDVEYLSEKOK
EYQWFLKSSDSLSLPEIPRANGIYIEBELIPVRENHGLSYSESPILLFPANOV
KERFLHLEKRYLSLFEKRSREVLCAPEKLLSFEERKVLSEISIGSLVGLPENN
LVSILKEAPSFKGIKREVRKISIELRAGKMKIVLTSFQAQTRLOGLKEKGVILL
NEDETEPLPHLGNHKSDFLIVSELNGLFLENOKLILISENDIFGEYKARKFPK
ONSALQSFIDIKGQDVVHIIHGVGKFLKERTSAGKERDFLKLRYSGGDSLFPPL
DOISLVORVIGCTSPRLDSLIGKSTWKTGDKYDQAVBALAEDLVQVYNSNKLQGYA
PPDITTYOESEBAFEYEETPDQIEALBAVKKDLSSRPMDRLVCGVGVGKTVAIR
AARVAMAGROIMMLAPTTILALQHVNTFKNRENYVRVAVELVSRFKTPAIRIDILAD

FSAGKVDVWVGTTHAIISSKPKNIGILIIIDEPORFGVNHKETIKKPNVYDUTLTA
TPIPRTLMALTGIRELSITATPPKNRSVTTYVLEEDDILISAIIRNEIORGQVFX

Query Match 83.2% Score 20.8; DB 1; Length 13356;
Best Local Similarity 91.7% Pred. No. 1.3e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTTCAGAT 24
DB 1339 AAAAAACAAAGATCCATTTCAGAT 1316

RESULT 27

LOCUS CO609503 31271 bp. DNA linear PAT 02-FEB-2004
DEFINITION Sequence 37261 from Patent WO0171042.
ACCESSION CO609503
VERSION CO609503.1 GI:41661648
KEYWORDS
SOURCE
ORGANISM Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
Patent: WO 0171042-A 37261 27-SEP-2001;
PE Corporation (NY) (US)

FEATURES

source location/Qualifiers
1..31271
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN

Query Match 83.2% Score 20.8; DB 6; Length 31271;
Best Local Similarity 91.7% Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTTCAGAT 24
DB 6323 AAAAAAATGATTCATTTCAGAT 6346

RESULT 28

LOCUS BX510318/c 58041 bp. DNA linear ROD 17-JUN-2003
DEFINITION Mouse DNA sequence from clone RP23-2901 on chromosome 4, complete
sequence.
ACCESSION BX510318
VERSION BX510318.4 GI:31873525
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 58041)

REFERENCE

1 Ellington, A.
Direct Submission
Submitted (17-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 17, 2003 this sequence version replaced gi:3141865
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC

Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-2901 is
constructed by the group of Pieter de Jong.
For further details see <http://www.choiti.org/bacpac/home.htm>
VECTOR: pBACE3.6.

FEATURES

source

location/Qualifiers
1..58041
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-2901"
/clone_1fb="RP23-2901"

ORIGIN

Query Match 83.2% Score 20.8; DB 9; Length 58041;
Best Local Similarity 91.7% Pred. No. 8.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTTCAGAT 24
DB 22064 AAAAAAAGTTCATTTCAGAT 22041

RESULT 29

LOCUS AL162395/c 62860 bp. DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-62L10 on chromosome 9 Contains a
ribosomal protein L24 (RPL24) pseudogene, complete sequence.
ACCESSION AL162395
VERSION AL162395.15 GI:16972811
KEYWORDS HTG; RPL24.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 62860)

REFERENCE

1 Skuce, C.
Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:15131472.

COMMENT

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr9
RP11-62L10 is from the library RP11-11.1 constructed by the group
of Pletier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: PBAC3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.

FEATURES

source

1..62860
/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-62L10"

/clone_11b="RP11-11.1"

/note="Clone_left_end: RP11-62L10"

16471
/note="Clone_right_end: RP11-418K10"

26383..26804
/locus_tag="RP11-62L10.1-001"

/locus_tag="RP11-62L10.1-001"

26383..26804
/locus_tag="RP11-62L10.1-001"

/note="match: proteins: P38663"

/pseudo

/codon_start=1

60861

/note="Clone_left_end: RP11-80H12"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 62860;
Best Local Similarity 91.7%; Pred. No. 7.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY 1 AAAAAAAAAAGTTCATTCAGAT 24
|||||

Db 34383 AAAAAAAAAAGTTCATTCAGAT 34360
|||||

RESULT 30
AP000243 96174 bp DNA linear PRI 20-NOV-1999

LOCUS AP000243 Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AMU region,
DEFINITION clone:H335C24, complete sequence.

ACCESSION AP000243

VERSION AP000243.1 GI:4835612

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Homo.

1 (bases 1 to 96174)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

Homosapiens 96,174bp genomic DNA of 21q22.1

Published Only in Database (1999)

2 (bases 1 to 96174)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

Direct Submission

JOURNAL

Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)

COMMENT

The sequence is a part of the data (ACCESSION No. AP000203 -
AP000214).
The sequencing project is supported by Japan Science Technology
Corporation (JST) and The Institute of Physical and Chemical
Research (RIKEN).

FEATURES

source

Location/Qualifiers

1..96174
/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

/clone="B335C24"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 96174;
Best Local Similarity 91.7%; Pred. No. 6.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY 1 AAAAAAAAAAGTTCATTCAGAT 24
|||||

Db 46120 AAAAAAAAAAGTTCATTCAGAT 46143
|||||

RESULT 31
AP000203 100000 bp DNA linear PRI 08-JAN-2000

LOCUS AP000203 Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AMU region,
DEFINITION clone f43D11-119B8, segment 1/12, complete sequence.

ACCESSION AP000203

VERSION AP000203.1 GI:4827141

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Homidae; Homo.

1 (bases 1 to 100000)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

Homosapiens 1,109,292bp genomic DNA of 21q22.1 (REGION:
D21S226-AMU CLONE RANGE: f43D11-119B8)

Published Only in Database (1999)

2 (bases 1 to 100000)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.

Direct Submission

Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)

E. coli transposon insertion: The present data does not contain E.
coli transposon sequences which integrated in the
original/previous sequences. We determined the boundary between
the insertion and genomic sequences experimentally, removed the
insertion sequences, reconstituted the present data. The sequencing
project is supported by Japan Science Technology Corporation (JST)

and The Institute of Physical and Chemical Research (RIKEN).

Location/Qualifiers

1..100000
/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q22.1"

```

ORIGIN
Query Match      83.2%  Score 20.8; DB 8; Length 100000;
Best Local Similarity 91.7%  Pred. No. 6.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTGAT 24
    |||||
Db 67150 AAAAAAAAAAGTTCATTGAT 67173

RESULT 32
AC020331/c 106659 bp DNA linear HTG 03-JAN-2000
LOCUS AC020331
DEFINITION AC020331
ACCESSION AC020331
VERSION AC020331.1 GI:6664566
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Archipoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
          1 (bases 1 to 106659)
REFERENCE Adams, M. and Venter, J.C.
AUTHORS Direct Submission
TITLE Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD, USA
COMMENT This sequence was identified as CDM.10212797 by the submitter.
          For more information on this record e-mail to fly@celera.com.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.

FEATURES
source
    1..106659
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
    /db_xref="taxon:7227"

ORIGIN
Query Match      83.2%  Score 20.8; DB 14; Length 106659;
Best Local Similarity 91.7%  Pred. No. 6.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTGAT 24
    |||||
Db 58527 AAAAAAAAAAGTTCATTGAT 58504

RESULT 33
AC153420_0/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC153420 Accession AC153420
Fragment Name Begin End
AC153420_0 1 110000
AC153420_1 100001 210000
AC153420_2 200001 310000
AC153420_3 300001 398932
LOCUS AC153420 398932 bp DNA linear HTG 14-DEC-2004
DEFINITION Mus musculus clone RP23-181K7, WORKING DRAFT SEQUENCE, 36 unordered
pieces
ACCESSION AC153420 AC124647
VERSION AC153420.4 GI:56565609
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Murioidea; Muridae; Murinae; Mus.
          1 (bases 1 to 398932)
REFERENCE Muray, D., Adams, C., Agbai II, O., Allen, C., Albrooks, S., Archer, P.,
AUTHORS Arredondo, H., Bandaranaike, D., Bangura, L., Beltran, B., Beltran, R.,

```

```

Berarducci, A., Biswal, K., Bjeth, P., Bonham, H., Buhay, C., Burch, P.,
Cadore, I., Canada, A., Cardenas, V., Carter, K., Cavazos, I.,
Chacko, J., Chan, J., Chavaz, D., Chen, A., Chen, G., Chen, R.,
Cheng, M.-T., Chin, J., Clerc, K., Cockrell, R., Coyle, M., Cree, A.,
Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De
Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H.,
Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K.,
Dzulanda, D., Egan, A., Escotto, M., Espinosa, V., Eugene, C., Fa, M.,
Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P.,
Fowler, G., Fu, Q., Fuh, E., Garcia, A., Garcia, R., Garner, T.,
Gaskin, C., Gench, S., Ghose, S., Gill, R., Gonzalez, D.,
Gonzalez-Garay, M., Guevara, W., Holder, M., Hailand, W., Heberlein, K.,
Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harrie, R., Haylak, P.,
Hawes, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, B.,
Hines, S., Hitchens, M., Hodgson, A., Hognes, M., Hollins, B.,
Howell, L., T., Huylk, S., Hume, J., Imo, K., Jackson, A., Jackson, L.,
Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kalafatis, K., Kelly, S.,
Keys, T., Khan, Z., King, L., Kovar, C., Kowis, A., Kowis, C., Lara, P.,
Leal, S., Lee, K., Lee, S., Legall, F.I., Lemon, S., Lewis, L., Li, B.,
Li, Y., Li, Z., Linell, M., Liu, W., Liu, Y.-S., Liu, Y., Liyange, D.,
London, P., Lopez, J., Lorensuhewa, L., Lozano, R., Luk, T., Madu, R.,
Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E.,
McClelland, H., McPherson, J., Mercado, C., Metzger, M.,
Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Munidasa, M.,
Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E.,
Nott, A., Nwokediemeh, O., Obregon, M., Ochi-Okorie, C., Odeh, E.,
Okwuon, G., Okwuon, K., Parker, D., Pasternak, S., Patel, B.,
Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T.,
Primus, E., Pu, L.-L., Puazo, M., Qin, X., Quinn, A., Quiroz, J.,
Rabara, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S.,
Rives, C., Rodriguez, F., Rojas, A., Ruiz, S.J., Sam, M., Sanders, M.,
Santhanasu, V., Santos, R., Savery, G., Scherer, S., Shen, H., Shen, Y.,
Sisson, I., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R.,
Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thörn, R., Thornton, R.,
Trejos, Z., Usmami, K., Varago, C., Verduzo, D., Villanueva, D., Virk, D.,
Volkov, X., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J.,
Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R.,
Wu, J., Yakub, S., Yan, K., Yan, Y., Yu, F., Zhang, J., Zhang, L.,
Zhang, Z., Zhou, J., Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 398932)
REFERENCE
AUTHORS Morley, K.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2004) Human Genome Sequencing Center, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 398932)
REFERENCE
AUTHORS Morley, K.C.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-2004) Human Genome Sequencing Center, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 14, 2004 this sequence version replaced gi:56553097.

COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help.tmc.edu
----- Project Information
Center project name: MBOR
Center clone name: RP23-181K7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 446510 bases at least Q40
Consensus quality: 446869 bases at least Q40
Consensus quality: 453002 bases at least Q20
Estimated insert size: 464853; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

```

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 36 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 7743: contig of 7743 bp in length
7744 7843: gap of unknown length
7844 19700: contig of 11857 bp in length
19701 19800: gap of unknown length
19801 40982: contig of 21182 bp in length
40983 41082: gap of unknown length
41083 59772: contig of 18690 bp in length
59773 59872: gap of unknown length
59873 90949: contig of 31077 bp in length
90950 91049: gap of unknown length
91050 11318: contig of 22269 bp in length
11319 113418: gap of unknown length
113419 130445: contig of 16927 bp in length
130446 130445: gap of unknown length
130446 163468: contig of 33023 bp in length
163469 163568: gap of unknown length
163569 194647: contig of 31079 bp in length
194648 223030: contig of 28283 bp in length
223031 223130: gap of unknown length
223131 225176: contig of 2046 bp in length
225177 225276: gap of unknown length
225277 228304: contig of 3028 bp in length
228305 228404: gap of unknown length
228405 231501: contig of 3097 bp in length
231502 231601: gap of unknown length
231602 233805: contig of 2204 bp in length
233806 233905: gap of unknown length
233906 237109: contig of 3204 bp in length
237110 237209: gap of unknown length
237210 242300: contig of 5091 bp in length
242301 242400: gap of unknown length
242401 247046: contig of 4646 bp in length
247047 247146: gap of unknown length
247147 255057: contig of 7911 bp in length
255058 255157: gap of unknown length
255158 260962: contig of 5805 bp in length
260963 261062: gap of unknown length
261063 308367: contig of 47305 bp in length
308368 308467: gap of unknown length
308468 317135: contig of 8668 bp in length
317136 317235: gap of unknown length
317236 328002: contig of 10767 bp in length
328003 328102: gap of unknown length
328103 338307: contig of 10205 bp in length
338308 338407: gap of unknown length
338408 349206: contig of 10799 bp in length
349207 349306: gap of unknown length
349307 356422: contig of 7116 bp in length
356423 356522: gap of unknown length
356523 356825: contig of 9303 bp in length
356826 365825: gap of unknown length
365827 368725: contig of 2800 bp in length
368726 368825: gap of unknown length
368826 371529: contig of 2704 bp in length
371530 371629: gap of unknown length
371630 374073: contig of 2444 bp in length
374074 374173: gap of unknown length
374174 376401: contig of 2228 bp in length
376402 376501: gap of unknown length
376502 379276: contig of 2775 bp in length
379277 379376: gap of unknown length
379377 383108: contig of 3732 bp in length
383109 383208: gap of unknown length
383209 387646: contig of 4438 bp in length

```

```

* 387647 387746: gap of unknown length
* 387747 391101: contig of 3355 bp in length
* 391102 391201: gap of unknown length
* 391202 395686: contig of 4485 bp in length
* 395687 395786: gap of unknown length
* 395787 398932: contig of 3146 bp in length.

```

FEATURES
 source Location/Qualifier

```

1. 398932
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /clone="RP23-181K7"
7744..7843
   /estimated_length=unknown
19701..19800
   /estimated_length=unknown
40983..41082
   /estimated_length=unknown
59773..59872
   /estimated_length=unknown
90950..91049
   /estimated_length=unknown
11319..113418
   /estimated_length=unknown
130446..130445
   /estimated_length=unknown
163469..163568
   /estimated_length=unknown
194648..194747
   /estimated_length=unknown
223031..223130
   /estimated_length=unknown
225177..225276

```

```

Query Match 83.2%; Score 20.8; DB 14; Length 110000;
Best Local Similarity 91.7%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
    |||||
Db 696 AAAAAATAAAGTTCATTCAGAT 673

```

```

RESULT 34
AC162929 1/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC162929 Accession AC162929
Fragment Name Begin End
AC162929_0 1 110000
AC162929_1 100001 210000
AC162929_2 200001 310000
AC162929_3 300001 392891
Continuation (2 of 4) of AC162929 from base 100001 (AC162929 Mus musculus chromosome 9)

```

```

Query Match 83.2%; Score 20.8; DB 14; Length 110000;
Best Local Similarity 91.7%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
    |||||
Db 58561 AAAAAAAGTCCAGTTCAGAT 58538

```

```

RESULT 35
BX640582 1
WPCOMMENT
Sequence split into 6 fragments LOCUS BX640582 Accession BX640582
Fragment Name Begin End
BX640582_0 1 110000
BX640582_1 100001 210000
BX640582_2 200001 310000
BX640582_3 300001 410000
BX640582_4 400001 510000

```

BK640582_5 500001 520210
Continuation (2 of 6) of BK640582 from base 100001 (BK640582 Danio rerio clone DKRY-1611)

Query Match 83.2%; Score 20.8; DB 14; Length 110000;
Best Local Similarity 91.7%; Pred. No. 6.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
DB 84923 AAAAAAAAAAGTTCATTCAGAT 84946

RESULT 36
AP008214_099/c
WPCOMMENT

Sequence split into 285 fragments LOCUS AP008214 Accession AP008214

Fragment Name	Begin	End
AP008214_000	1	110000
AP008214_001	100001	210000
AP008214_002	200001	310000
AP008214_003	300001	410000
AP008214_004	400001	510000
AP008214_005	500001	610000
AP008214_006	600001	710000
AP008214_007	700001	810000
AP008214_008	800001	910000
AP008214_009	900001	1010000
AP008214_010	1000001	1110000
AP008214_011	1100001	1210000
AP008214_012	1200001	1310000
AP008214_013	1300001	1410000
AP008214_014	1400001	1510000
AP008214_015	1500001	1610000
AP008214_016	1600001	1710000
AP008214_017	1700001	1810000
AP008214_018	1800001	1910000
AP008214_019	1900001	2010000
AP008214_020	2000001	2110000
AP008214_021	2100001	2210000
AP008214_022	2200001	2310000
AP008214_023	2300001	2410000
AP008214_024	2400001	2510000
AP008214_025	2500001	2610000
AP008214_026	2600001	2710000
AP008214_027	2700001	2810000
AP008214_028	2800001	2910000
AP008214_029	2900001	3010000
AP008214_030	3000001	3110000
AP008214_031	3100001	3210000
AP008214_032	3200001	3310000
AP008214_033	3300001	3410000
AP008214_034	3400001	3510000
AP008214_035	3500001	3610000
AP008214_036	3600001	3710000
AP008214_037	3700001	3810000
AP008214_038	3800001	3910000
AP008214_039	3900001	4010000
AP008214_040	4000001	4110000
AP008214_041	4100001	4210000
AP008214_042	4200001	4310000
AP008214_043	4300001	4410000
AP008214_044	4400001	4510000
AP008214_045	4500001	4610000
AP008214_046	4600001	4710000
AP008214_047	4700001	4810000
AP008214_048	4800001	4910000
AP008214_049	4900001	5010000
AP008214_050	5000001	5110000
AP008214_051	5100001	5210000
AP008214_052	5200001	5310000
AP008214_053	5300001	5410000
AP008214_054	5400001	5510000
AP008214_055	5500001	5610000

AP008214_056	5600001	5710000
AP008214_057	5700001	5810000
AP008214_058	5800001	5910000
AP008214_059	5900001	6010000
AP008214_060	6000001	6110000
AP008214_061	6100001	6210000
AP008214_062	6200001	6310000
AP008214_063	6300001	6410000
AP008214_064	6400001	6510000
AP008214_065	6500001	6610000
AP008214_066	6600001	6710000
AP008214_067	6700001	6810000
AP008214_068	6800001	6910000
AP008214_069	6900001	7010000
AP008214_070	7000001	7110000
AP008214_071	7100001	7210000
AP008214_072	7200001	7310000
AP008214_073	7300001	7410000
AP008214_074	7400001	7510000
AP008214_075	7500001	7610000
AP008214_076	7600001	7710000
AP008214_077	7700001	7810000
AP008214_078	7800001	7910000
AP008214_079	7900001	8010000
AP008214_080	8000001	8110000
AP008214_081	8100001	8210000
AP008214_082	8200001	8310000
AP008214_083	8300001	8410000
AP008214_084	8400001	8510000
AP008214_085	8500001	8610000
AP008214_086	8600001	8710000
AP008214_087	8700001	8810000
AP008214_088	8800001	8910000
AP008214_089	8900001	9010000
AP008214_090	9000001	9110000
AP008214_091	9100001	9210000
AP008214_092	9200001	9310000
AP008214_093	9300001	9410000
AP008214_094	9400001	9510000
AP008214_095	9500001	9610000
AP008214_096	9600001	9710000
AP008214_097	9700001	9810000
AP008214_098	9800001	9910000
AP008214_099	9900001	10010000
AP008214_100	10000001	10110000
AP008214_101	10100001	10210000
AP008214_102	10200001	10310000
AP008214_103	10300001	10410000
AP008214_104	10400001	10510000
AP008214_105	10500001	10610000
AP008214_106	10600001	10710000
AP008214_107	10700001	10810000
AP008214_108	10800001	10910000
AP008214_109	10900001	11010000
AP008214_110	11000001	11110000
AP008214_111	11100001	11210000
AP008214_112	11200001	11310000
AP008214_113	11300001	11410000
AP008214_114	11400001	11510000
AP008214_115	11500001	11610000
AP008214_116	11600001	11710000
AP008214_117	11700001	11810000
AP008214_118	11800001	11910000
AP008214_119	11900001	12010000
AP008214_120	12000001	12110000
AP008214_121	12100001	12210000
AP008214_122	12200001	12310000
AP008214_123	12300001	12410000
AP008214_124	12400001	12510000
AP008214_125	12500001	12610000
AP008214_126	12600001	12710000
AP008214_127	12700001	12810000
AP008214_128	12800001	12910000

AP008214_129 12900001 13010000
 AP008214_130 13000001 13110000
 AP008214_131 13100001 13210000
 AP008214_132 13200001 13310000
 AP008214_133 13300001 13410000
 AP008214_134 13400001 13510000
 AP008214_135 13500001 13610000
 AP008214_136 13600001 13710000
 AP008214_137 13700001 13810000
 AP008214_138 13800001 13910000
 AP008214_139 13900001 14010000
 AP008214_140 14000001 14110000
 AP008214_141 14100001 14210000
 AP008214_142 14200001 14310000
 AP008214_143 14300001 14410000
 AP008214_144 14400001 14510000
 AP008214_145 14500001 14610000
 AP008214_146 14600001 14710000
 AP008214_147 14700001 14810000
 AP008214_148 14800001 14910000
 AP008214_149 14900001 15010000
 AP008214_150 15000001 15110000
 AP008214_151 15100001 15210000
 AP008214_152 15200001 15310000
 AP008214_153 15300001 15410000
 AP008214_154 15400001 15510000
 AP008214_155 15500001 15610000
 AP008214_156 15600001 15710000
 AP008214_157 15700001 15810000
 AP008214_158 15800001 15910000
 AP008214_159 15900001 16010000
 AP008214_160 16000001 16110000
 AP008214_161 16100001 16210000
 AP008214_162 16200001 16310000
 AP008214_163 16300001 16410000
 AP008214_164 16400001 16510000
 AP008214_165 16500001 16610000
 AP008214_166 16600001 16710000
 AP008214_167 16700001 16810000
 AP008214_168 16800001 16910000
 AP008214_169 16900001 17010000
 AP008214_170 17000001 17110000
 AP008214_171 17100001 17210000
 AP008214_172 17200001 17310000
 AP008214_173 17300001 17410000
 AP008214_174 17400001 17510000
 AP008214_175 17500001 17610000
 AP008214_176 17600001 17710000
 AP008214_177 17700001 17810000
 AP008214_178 17800001 17910000
 AP008214_179 17900001 18010000
 AP008214_180 18000001 18110000
 AP008214_181 18100001 18210000
 AP008214_182 18200001 18310000
 AP008214_183 18300001 18410000
 AP008214_184 18400001 18510000
 AP008214_185 18500001 18610000
 AP008214_186 18600001 18710000
 AP008214_187 18700001 18810000
 AP008214_188 18800001 18910000
 AP008214_189 18900001 19010000
 AP008214_190 19000001 19110000
 AP008214_191 19100001 19210000
 AP008214_192 19200001 19310000
 AP008214_193 19300001 19410000
 AP008214_194 19400001 19510000
 AP008214_195 19500001 19610000
 AP008214_196 19600001 19710000

Query Match 83.2% Score 20.8; DB 15; Length 110000;
 Best Local Similarity 91.7% Pred. No. 6.6e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTTCGATTCAGATA 25
 Db 12370 AAAAAAAAAAGTTCGATTCATCATATA 12347
 RESULT 37
 AC124958
 LOCUS
 DEFINITION Medicago truncatula clone mch2-18111, complete sequence.
 AC124958
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Medicago truncatula (barrel medic)
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 1 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Medicago truncatula BAC Clone mch2-18111
 Unpublished
 2 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Direct Submission
 Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 3 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Direct Submission
 Submitted (04-MAR-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 4 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Direct Submission
 Submitted (12-NOV-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 5 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Direct Submission
 Submitted (04-MAR-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 6 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Direct Submission
 Submitted (05-MAR-2003) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 7 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Direct Submission
 Submitted (24-JUN-2004) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 8 (bases 1 to 120562)
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
 Cook,D., Kim,D. and Roe,B.A.
 Direct Submission
 Submitted (08-JAN-2005) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
 On Jun 24, 2004 this sequence version replaced g1:45267932.

```

----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
  source
    1. 120562
       /organism="Medicago truncatula"
       /mol_type="genomic DNA"
       /db_xref="taxon:3880"
       /clone="mth2-1811"
       /clone_lib="Medicago truncatula BAC library H2"

ORIGIN

Query Match      83.2%; Score 20.8; DB 15; Length 120562;
Best Local Similarity 91.7%; Pred. No. 6.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      2 AAAAAAAAAAGTTCATTCAGATA 25
Db      21342 AAAAAAGAAATCCATTCAGATA 21365

RESULT 38
AC144474/c 142202 bp DNA linear PLN 09-APR-2005
LOCUS      AC144474
DEFINITION Medicago truncatula clone mth2-1515, complete sequence.
ACCESSION  AC144474
VERSION     AC144474.22 GI:62460761
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
ORGANISM    Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosoids; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
            Medicago.
            1 (bases 1 to 142202)
            Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
            Medicago truncatula BAC Clone mth2-1515
            2 (bases 1 to 142202)
            Unpublished
            3 (bases 1 to 142202)
            Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
            Direct Submission
            Submitted (22-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            4 (bases 1 to 142202)
            Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
            Direct Submission
            Submitted (16-NOV-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            5 (bases 1 to 142202)
            Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
            Direct Submission
            Submitted (30-NOV-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            6 (bases 1 to 142202)
            Submitted (26-FEB-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            7 (bases 1 to 142202)
            Submitted (09-APR-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
            On Apr 9, 2005 this sequence version replaced gi:61675768.

----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

```

```

JOURNAL
Submitted (22-MAR-2005) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
7 (bases 1 to 142202)
Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (09-APR-2005) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Apr 9, 2005 this sequence version replaced gi:61675768.

COMMENT
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
  source
    1. 142202
       /organism="Medicago truncatula"
       /mol_type="genomic DNA"
       /db_xref="taxon:3880"
       /clone="mth2-1515"
       /clone_lib="Medicago truncatula BAC library H2"
       /note="This is one of two clones in the same well from
       mth2-1515"

ORIGIN

Query Match      83.2%; Score 20.8; DB 15; Length 142202;
Best Local Similarity 91.7%; Pred. No. 6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      2 AAAAAAAAAAGTTCATTCAGATA 25
Db      99221 AAAAAAGAAATCCATTCAGATA 99198

RESULT 39
AC063921/c 144833 bp DNA linear HTG 08-JAN-2003
LOCUS      AC063921
DEFINITION Homo sapiens chromosome 3 clone RP11-142P10, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION  AC063921
VERSION     AC063921.18 GI:20335587
KEYWORDS    HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominoidea; Homo.
            1 (bases 1 to 144833)
            Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alibekova,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
            Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,B., Brown,M., Bryant,N.P.,
            Buhay,C., Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Derr,A.L., Ding,Y., Dinh,H.H.,
            Douchwaile,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gali,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
            Homsa,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W., Loulgesed,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

```

Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Messey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Mezzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogund, M., Okunnu, G.,
 Otagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shochet, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swalek, A., Tabors, P., Tameria, A., Tameria, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Uemami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.

 Title: Direct Submission
 Unpublished
 2 (bases 1 to 144833)
 Worley, K.C.

 Title: Direct Submission
 Submitted (22-APR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 144833)
 Worley, K.C.

 Title: Direct Submission
 Submitted (08-JUN-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 28, 2002 this sequence version replaced gi:16117924.

 Title: Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu

 Title: Project Information
 Center project name: HAXY
 Center clone name: RP11-142P10

 Title: Summary Statistics
 Sequencing vector: M13;
 Chemistry: Dye-Primer Bodypy: 14% of reads
 Chemistry: Dye-terminator Big Dye: 86% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 127260 bases at least Q40
 Consensus quality: 132735 bases at least Q30
 Consensus quality: 136183 bases at least Q20
 Estimated insert size: 139078; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html).
 * NOTE: This is a working draft sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 2437: contig of 2437 bp in length
 * 2438 2537: gap of unknown length
 * 2538 4812: contig of 2275 bp in length
 * 4813 4912: gap of unknown length
 * 4913 8488: contig of 3576 bp in length
 * 8489 8589: gap of unknown length
 * 8589 11163: contig of 2575 bp in length
 * 11164 11263: gap of unknown length
 * 11263 16638: contig of 5375 bp in length
 * 16639 16738: gap of unknown length
 * 16739 21870: contig of 5132 bp in length
 * 21871 21970: gap of unknown length

FEATURES
 source
 * 21971 30780: contig of 8810 bp in length
 * 30781 30880: gap of unknown length
 * 30881 39267: contig of 8387 bp in length
 * 39268 39368: gap of unknown length
 * 39368 46890: contig of 7523 bp in length
 * 46891 46990: gap of unknown length
 * 46991 59656: contig of 12666 bp in length
 * 59657 59757: gap of unknown length
 * 59757 83309: contig of 23553 bp in length
 * 83310 83410: gap of unknown length
 * 83410 111857: contig of 28448 bp in length
 * 111858 111957: gap of unknown length
 * 111958 144833: contig of 32876 bp in length.
 * Location/Qualifiers
 1. 144833
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-142P10"
 2438..2537
 /estimated_length=unknown
 4813..4912
 /estimated_length=unknown
 8489..8588
 /estimated_length=unknown
 11164..11263
 /estimated_length=unknown
 16639..16738
 /estimated_length=unknown
 21871..21970
 /estimated_length=unknown
 30781..30880
 /estimated_length=unknown
 39268..39367
 /estimated_length=unknown
 46891..46990
 /estimated_length=unknown
 59657..59756
 /estimated_length=unknown
 83310..83409
 /estimated_length=unknown
 111858..111957
 /estimated_length=unknown

 ORIGIN
 Query Match 83.2% Score 20.8; DB 14; Length 144833;
 Best Local Similarity 91.7% Pred. No. 6e+02; 2; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 2;
 QY 1 AAAAAAAAAAGTTCCAAATTCAGAT 24
 DB 65944 AAAAAAAAAAGTTCCCACTCAGAT 65921
 |||||
 RESULT 40 145633 bp DNA linear PRI 12-JUN-2004
 BS000107
 LOCUS Pan troglodytes chromosome 22 clone:PTB-020H02, map 22, complete
 DEFINITION
 sequences.
 ACCESSION BS000107.1 GI:37537374
 VERSION BS000107.1 GI:37537374
 KEYWORDS
 SOURCE
 ORGANISM
 Pan troglodytes (chimpanzee)
 Pan troglodytes
 Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominoidea; Pan.

 REFERENCE
 1 The International Chimpanzee Chromosome 22 Consortium.
 DNA sequence and comparative analysis of chimpanzee chromosome 22
 JOURNAL Nature 429, 382-388 (2004)
 REFERENCE 2 (bases 1 to 145633)

AUTHORS Hattori, M., Toyoda, A., Watanabe, H., Taylor, T. D., Kuroki, Y., Fujiyama, A. and Sakaki, Y.

TITLE Direct Submission

JOURNAL Submitted (12-MAY-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT

The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
 *Chinese National Human Genome Center at Shanghai, Shanghai, China;
 *GfF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute of Molecular Biotechnology, Jena, Germany; *KRIIBB Genome Research Center, Daejeon, Korea;
 *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
 *National Institute of Genetics, Mishima, Japan;
 *National Yang Ming University Genome Research Center, Taipei, Taiwan;
 *RIKEN Genomic Sciences Center, Yokohama, Japan.

Center: RIKEN Genomic Sciences Center
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp

Center project name: The Chimpanzee Chromosome 22 Sequencing Project
 Center clone name: PTB-020H02

Sequencing vector: pUC18, pUC13, pTZ19; 100% of reads Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly program: Phrap; version 0.990329

Consensus quality: 145,593 bases at least Q40
 Consensus quality: 36 bases at least Q30
 Consensus quality: 4 bases at least Q20

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Source information:
 The RCI-43 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at
 http://www.chori.org/bacpac/mechimp3.htm.
 The clone may be obtained from Pieter J. de Jong and coworkers (http://www.chori.org/bacpac).

VECTOR: pBACe3.6
 The CHORI-251 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at
 http://www.chori.org/bacpac/chimpanzee251.htm.
 The clone may be obtained from Pieter J. de Jong and coworkers (http://www.chori.org/bacpac).

VECTOR: pFRABAC2.1
 The PFI1 chimpanzee BAC library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.
 Clones may be obtained from Aaso Fujiyama and co-workers (http://www.gsc.riken.go.jp).
VECTOR: pKSI45

The PTF22 chimpanzee Fosmid library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.
 Clones may be obtained from Aaso Fujiyama and co-workers (http://www.gsc.riken.go.jp).
VECTOR: pKSI43

Sequence Quality Assessment:
 This entry has been annotated with sequence estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

FEATURES**source**

Neighboring clones: RP43-184113(left) and PTB-237C02(right).
 Location/Qualifiers
 1. 145633
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="22"
 /clone="PTB-020H02"
 /clone_1fb="PTB1 chimpanzee BAC"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 145633;
 Best Local Similarity 91.7%; Pred. No. 6e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 AAAAAAAAAAGTCCATTCCAGAT 24
 Db 38004 AAAAAAAAAAGTCCATTCCAGAT 38027

RESULT 41**LOCUS**

AC021981 149800 bp DNA linear HTG 19-JUL-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-11C15 map 11, WORKING DRAFT

ACCESSION

AC021981
 AC021981.2 GI:9280747
 HTG: HTGS PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE
 ORGANISM

REFERENCE**AUTHORS****TITLE****JOURNAL****REFERENCE****AUTHORS****TITLE****JOURNAL****REFERENCE****AUTHORS****TITLE****JOURNAL****REFERENCE****AUTHORS****TITLE****JOURNAL****REFERENCE**

1 (bases 1 to 149800)
 Birren, B., Linton, L., Nuebaum, C. and Lander, E.
 Homo sapiens chromosome 11, clone RP11-11C15
 Unpublished
 2 (bases 1 to 149800)
 Anderson, S., Baldwin, J., Barra, N., Beckerly, R., Beda, F., Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cooke, P., D'Arrellano, K., Dewar, K., Domini, M., Doyle, M., Feneceor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, J., Leoczky, J., Levine, R., Liu, C., Liu, G., Locke, K., McDougal, P., Margulis, N., McEwan, P., McGuck, A., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (23-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 19, 2000 this sequence version replaced gi:6731265.

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1335

Center clone name: 11_C15

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 13517 bases at least Q40

Consensus quality: 14404 bases at least Q30

Insert size: 12100; agarose-fp

Insert size: 14700; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 3.5 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 consists of 29 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 1441: contig of 1441 bp in length
* 1442 1541: gap of 100 bp
* 1542 2883: contig of 1342 bp in length
* 2884 2984: gap of 100 bp
* 2984 4649: contig of 1666 bp in length
* 4650 4749: gap of 100 bp
* 4750 6112: contig of 1363 bp in length
* 6113 6212: gap of 100 bp
* 6213 7413: contig of 1201 bp in length
* 7414 7513: gap of 100 bp
* 7514 9050: contig of 1537 bp in length
* 9051 9150: gap of 100 bp
* 9151 10944: contig of 1794 bp in length
* 10945 11044: gap of 100 bp
* 11045 13203: contig of 2159 bp in length
* 13204 13303: gap of 100 bp
* 13304 16118: contig of 2815 bp in length
* 16119 16218: gap of 100 bp
* 16219 18673: contig of 2455 bp in length
* 18674 18773: gap of 100 bp
* 18774 21180: contig of 2407 bp in length
* 21181 21280: gap of 100 bp
* 21281 23200: contig of 1920 bp in length
* 23201 23300: gap of 100 bp
* 23301 26526: contig of 3226 bp in length
* 26527 26626: gap of 100 bp
* 26627 29655: contig of 3029 bp in length
* 29656 29755: gap of 100 bp
* 29756 34738: contig of 4983 bp in length
* 34739 34838: gap of 100 bp
* 34839 39394: contig of 4556 bp in length
* 39395 39494: gap of 100 bp
* 39495 43669: contig of 4175 bp in length
* 43670 43769: gap of 100 bp
* 43770 49948: contig of 6179 bp in length
* 49949 50048: gap of 100 bp
* 50049 55578: contig of 5530 bp in length
* 55579 55678: gap of 100 bp
* 55679 60959: contig of 5281 bp in length
* 60960 61059: gap of 100 bp
* 61060 66394: contig of 5335 bp in length
* 66395 66494: gap of 100 bp

```

```

* 66495 72947: contig of 6453 bp in length
* 72948 73047: gap of 100 bp
* 73048 80874: contig of 7827 bp in length
* 80875 80974: gap of 100 bp
* 80975 87847: contig of 6873 bp in length
* 87848 87947: gap of 100 bp
* 87948 97279: contig of 9332 bp in length
* 97280 97379: gap of 100 bp
* 97380 106037: contig of 8658 bp in length
* 106038 106137: gap of 100 bp
* 106138 114207: contig of 8069 bp in length
* 114207 114306: gap of 100 bp
* 114307 129830: contig of 15524 bp in length
* 129831 129930: gap of 100 bp
* 129931 149800: contig of 19870 bp in length.

FEATURES
Source
1. .149800
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-11C15"
/clone_id="RP11 Human Male BAC"
1. .1441
/note="assembly_fragment"
1442. .1541
/estimated_length=100
1542. .2883
/note="assembly_fragment"
2884. .2983
/estimated_length=100
2984. .4649
/note="assembly_fragment"
4650. .4749
/estimated_length=100
4750. .6112
/note="assembly_fragment"
6113. .6212
/estimated_length=100
6213. .7413
/note="assembly_fragment"
7414. .7513
/estimated_length=100
7514. .9050
/note="assembly_fragment"
9051. .9150
/estimated_length=100
9151. .10944
/note="assembly_fragment"
10945. .11044
/estimated_length=100
11045. .13203
/note="assembly_fragment"
13204. .13303
/estimated_length=100
13304. .16118
/note="assembly_fragment"
16119. .16218
/estimated_length=100
16219. .18673
/note="assembly_fragment"
18674. .18773
/estimated_length=100
18774. .21180
/note="assembly_fragment"
21181. .21280
/estimated_length=100
21281. .23200
/note="assembly_fragment"
23201. .23300
gap
misc_feature
1441: contig of 1441 bp in length
1541: gap of 100 bp
2883: gap of 100 bp
2984: contig of 1666 bp in length
4749: gap of 100 bp
6112: contig of 1363 bp in length
6212: gap of 100 bp
7413: contig of 1201 bp in length
7513: gap of 100 bp
9050: contig of 1537 bp in length
9150: gap of 100 bp
10944: contig of 1794 bp in length
11044: gap of 100 bp
13203: contig of 2159 bp in length
13303: gap of 100 bp
16118: contig of 2815 bp in length
16218: gap of 100 bp
18673: contig of 2455 bp in length
18773: gap of 100 bp
21180: contig of 2407 bp in length
21280: gap of 100 bp
23200: contig of 1920 bp in length
23300: gap of 100 bp
26526: contig of 3226 bp in length
26626: gap of 100 bp
29655: contig of 3029 bp in length
29755: gap of 100 bp
34738: contig of 4983 bp in length
34838: gap of 100 bp
39394: contig of 4556 bp in length
39494: gap of 100 bp
43669: contig of 4175 bp in length
43769: gap of 100 bp
49948: contig of 6179 bp in length
50048: gap of 100 bp
55578: contig of 5530 bp in length
55678: gap of 100 bp
60959: contig of 5281 bp in length
61059: gap of 100 bp
66394: contig of 5335 bp in length
66494: gap of 100 bp

```

gene; rps8 gene; rpoC1 gene; rpoC2 gene; rps11 gene; rps12 gene;
-rps14 gene; rps15 gene; rps16 gene; rps18 gene; rps19 gene; rps2
gene; rps3 gene; rps4 gene; rps7 gene; rps8 gene; rns16 gene;
rnm23 gene; rnm4_5S gene; rns5 gene; transfer RNA-Ala (ucc) ;
transfer RNA-Arg (acg) ; transfer RNA-Arg (ucc) ; transfer RNA-Asn
(guu) ; transfer RNA-Asp (guc) ; transfer RNA-Cys (gca) ; transfer
RNA-Glu (uuu) ; transfer RNA-Gln (uug) ; transfer RNA-Ile (auc) ;
transfer RNA-Lys (cag) ; transfer RNA-Met (cau) ; transfer RNA-Phe
(gau) ; transfer RNA-Pro (ugu) ; transfer RNA-Ser (uca) ; transfer
RNA-Thr (agu) ; transfer RNA-Trp (cca) ; transfer RNA-Tyr (gua)
(guu) ; transfer RNA-Val (gac) ; transfer RNA-Val (uac) ; transfer
RNA-Leu (caa) ; transfer RNA-Ile (cau) ; transfer RNA-Ile (gau) ; transfer
transfer RNA-Lys (uuu) ; transfer RNA-Met (cau) ; transfer RNA-Phe
(gau) ; transfer RNA-Pro (ugu) ; transfer RNA-Ser (uca) ; transfer
RNA-Thr (agu) ; transfer RNA-Trp (cca) ; transfer RNA-Tyr (gua)
(guu) ; transfer RNA-Val (gac) ; transfer RNA-Val (uac) ; transfer
RNA-Ala (ucc) gene; tRNA-Arg (acg) gene; tRNA-Arg (ucc) gene; tRNA-Asn
(guu) gene; tRNA-Asp (guc) gene; tRNA-Cys (gca) gene; tRNA-Glu
(uuc) gene; tRNA-Gly (ggc) gene; tRNA-His (aug) gene; tRNA-Ile
(cau) gene; tRNA-Ile (gau) gene; tRNA-Leu (caa) gene; tRNA-Leu
(uua) gene; tRNA-Phe (gaa) gene; tRNA-Lys (uuu) gene; tRNA-Met
(cau) gene; tRNA-Ser (gga) gene; tRNA-Ser (uga) gene; tRNA-Thr
(gcu) gene; tRNA-Thr (ugu) gene; tRNA-Tip (cca) gene; tRNA-Tyr
(gua) gene; tRNA-Val (gac) gene; tRNA-Val (uac) gene; ycf1 gene;
ycf15 gene; ycf2 gene; ycf3 gene; ycf4 gene; ycf5 gene; ycf6 gene; ycf9
gene.

SOURCE
ORGANISM
chloroplast Calycanthus floridus var. glaucus
Calycanthus floridus var. glaucus
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; magnoliids; Laurales; Calycanthaceae;
Calycanthus.
1 (bases 1 to 15337)
Goremykin,V., Hirsch-Ernst,K., Wolf,S. and Hellwig,F.
Complete structure of the chloroplast genome of Calycanthus
floridis

JOURNAL
AUTHORS
TITLE
Unpublished
2 (bases 1 to 15337)
Goremykin,V.

JOURNAL
AUTHORS
TITLE
Direct Submission
Submitted (17-JAN-2002) Institut fuer Spezielle Botanik,
Friedrich-Schiller-Universitaet, Philosophenweg 16, Jena D-07743,
GERMANY

FEATURES
SOURCE
Location/Qualifiers
1..15337
/organism="Calycanthus floridus var. glaucus"
/organelle="plastid:chloroplast"
/mol_type="genomic DNA"
/variety="glaucus"
/db_xref="taxon:212734"
/note="aka var. ferax"
complement(join(4171..4207,1664..1698))
/gene="tRNA-Lys (uuu)"
complement(join(4171..4207,1664..1698))
/gene="tRNA-Lys (uuu)"
/product="tRNA-Lys"
/note="codon recognized: AAA"
complement(join(99138..99163,99700..99931,71034..71147)))
/note="trans-spliced"
/codon_start=1
/transl_table=11
/product="ribosomal protein S12"
/protein_id="CAD28682.1"
/db_xref="GI:32399426"
/db_xref="GOA:Q7Y677"
/db_xref="InterPro:IPR005679"
/db_xref="InterPro:IPR006032"
/db_xref="UniProt/TREMBL:Q7Y677"
translation="MTPIKILIRNTPRIIRNTKSPALRGCPORGTCTRYVTIPKK
PNSALKVARVRLTSFEITXIIPEIGSHNSQHSVVVLNGKVKDIPGVRIHIVRGTL
DAVGAKDRQQRSKTVKKPK"
join(complement(99700..99931),complement(99164..99699),
complement(99138..99163),complement(71034..71147)).

gene

```

CDS
140351..140582,141119..141144)
/gene="rps12"
join(complement(71034..71147),140351..140582,
141119..141144)
/gene="rps12"
/notes="trans-spliced"
/codon_start=1
/transl_table=11
/product="ribosomal protein S12"
/protein_id="CAD28781.1"
/db_xref="GI:32399440"
/db_xref="GOA:Q76777"
/db_xref="InterPro:IPR005679"
/db_xref="InterPro:IPR006032"
/db_xref="UniProt/TREMBL:Q76777"
/translacion="MPTIKOLIRNTROPINRTPSPALRGCPORGRCTRVYITTPK
PNSALRKVARVRLTSGREITAYIPGIGHNOGSHVVLVRGRYKDLPGVAYHIVRGTL
DAVGKDRQGRSKRYKKPK"
complement(1..78)
/gene="tRNA-His (gug)"
complement(1..78)
/gene="tRNA-His (gug)"
/product="tRNA-His"
/notes="codon recognized: CAC"
complement(366..1427)
/gene="psbA"
complement(366..1427)
/gene="psbA"
/codon_start=1
/transl_table=11
/product="PsbI 32 kD protein"
/protein_id="CAD28701.1"
/db_xref="GI:3239359"
/db_xref="GOA:Q7YJY8"
/db_xref="InterPro:IPR000484"
/db_xref="InterPro:IPR005867"
/db_xref="UniProt/TREMBL:Q7YJY8"
/translacion="MTAILRREBTSJMGRCWMTSTENRKYIGMFGVIMPTLTLTA
TSVPIIAFPAIPDIDIGIRPVSGLYGNLISGAIIPISAIGLHPYIMEASV
DEMLNGCPPELVIAHFLGVACYMGREMLSPFLGMRPMIAVYSAVAAATVFLI
YPIGQSGFSQDMPIGISTGTPMFIPOAEHNIIMHPHMGVAGVCGSGSLPSMHGSL
VTSILIRRTTENSANAGYRPGQEEETNTVAAGYGRLLIPGASNNRSRHLFPLA
AMPVGIWFTALGISTAFVNLNGFNQSVVDSGRVINTADIIINPANGMEVHER
NAHNPFLDLAAVEVPSNG"
complement(1664..1698)
/gene="rps12"
/product="transfer RNA-Lys (uuu)"
complement(1699..4170)
/gene="tRNA-Lys (uuu)"
complement(1699..4170)
/gene="tRNA-Lys (uuu)"
complement(1958..3475)
/gene="matK"
complement(1958..3475)
/gene="matK"
/notes="orf within trnK intron"
/codon_start=1
/transl_table=11
/product="maturase"
/protein_id="CAD28702.1"
/db_xref="GI:32399360"
/db_xref="GOA:Q7YJY7"
/db_xref="InterPro:IPR002866"
/db_xref="InterPro:IPR00442"
/db_xref="UniProt/TREMBL:Q7YJY7"
/translacion="MEELQGLYELIDGRQHFLYPLLQEVYIALADHGLNSILSE
HMENLSDNKSSSLIVKRLITRMQONHFIISVDSNQKGVGNKPKFSQKISGEGRA
VIVIPSLQVSSLEKEIKAFNNSRSHISIPFEDDKSHLHNVDDLIPVITHE
ILVOTLCWIDQADPSLHLRFLFHEYNNNSLITPKKISIFFSKENDRSLFLYNSHV
YECESVFIPLKQSHLRSTSGSFLERTHYGKIEHLVVVLGNDPXTLMLFDPDV
HYVYQKSLIASGTOFLIKKMYHLVNFQWQGYLMSQODRIHLNQLNHSYFPG
YLSVOLNNSVVRQMLENAFIMDTAIKKPFTVPIIPIIGSLAKAKKSGSGPIEK
PFRDLDSEIINFRKICKNLSHYHSGSSKKQSLYRIKPLRLSCARTLSRAKKSIV

```

```

exon
complement(4171..4207)
/gene="rps12"
complement(5012..6111)
/gene="rps16"
complement(join(5012..5238,6072..6111))
/gene="rps16"
/notes="trans-spliced"
/codon_start=1
/transl_table=11
/product="ribosomal protein S16"

CDS
140351..140582,141119..141144)
/gene="rps12"
join(complement(71034..71147),140351..140582,
141119..141144)
/gene="rps12"
/notes="trans-spliced"
/codon_start=1
/transl_table=11
/product="ribosomal protein S12"
/protein_id="CAD28781.1"
/db_xref="GI:32399440"
/db_xref="GOA:Q76777"
/db_xref="InterPro:IPR005679"
/db_xref="InterPro:IPR006032"
/db_xref="UniProt/TREMBL:Q76777"
/translacion="MPTIKOLIRNTROPINRTPSPALRGCPORGRCTRVYITTPK
PNSALRKVARVRLTSGREITAYIPGIGHNOGSHVVLVRGRYKDLPGVAYHIVRGTL
DAVGKDRQGRSKRYKKPK"
complement(1..78)
/gene="tRNA-His (gug)"
complement(1..78)
/gene="tRNA-His (gug)"
/product="tRNA-His"
/notes="codon recognized: CAC"
complement(366..1427)
/gene="psbA"
complement(366..1427)
/gene="psbA"
/codon_start=1
/transl_table=11
/product="PsbI 32 kD protein"
/protein_id="CAD28701.1"
/db_xref="GI:3239359"
/db_xref="GOA:Q7YJY8"
/db_xref="InterPro:IPR000484"
/db_xref="InterPro:IPR005867"
/db_xref="UniProt/TREMBL:Q7YJY8"
/translacion="MTAILRREBTSJMGRCWMTSTENRKYIGMFGVIMPTLTLTA
TSVPIIAFPAIPDIDIGIRPVSGLYGNLISGAIIPISAIGLHPYIMEASV
DEMLNGCPPELVIAHFLGVACYMGREMLSPFLGMRPMIAVYSAVAAATVFLI
YPIGQSGFSQDMPIGISTGTPMFIPOAEHNIIMHPHMGVAGVCGSGSLPSMHGSL
VTSILIRRTTENSANAGYRPGQEEETNTVAAGYGRLLIPGASNNRSRHLFPLA
AMPVGIWFTALGISTAFVNLNGFNQSVVDSGRVINTADIIINPANGMEVHER
NAHNPFLDLAAVEVPSNG"
complement(1664..1698)
/gene="rps12"
/product="transfer RNA-Lys (uuu)"
complement(1699..4170)
/gene="tRNA-Lys (uuu)"
complement(1699..4170)
/gene="tRNA-Lys (uuu)"
complement(1958..3475)
/gene="matK"
complement(1958..3475)
/gene="matK"
/notes="orf within trnK intron"
/codon_start=1
/transl_table=11
/product="maturase"
/protein_id="CAD28702.1"
/db_xref="GI:32399360"
/db_xref="GOA:Q7YJY7"
/db_xref="InterPro:IPR002866"
/db_xref="InterPro:IPR00442"
/db_xref="UniProt/TREMBL:Q7YJY7"
/translacion="MEELQGLYELIDGRQHFLYPLLQEVYIALADHGLNSILSE
HMENLSDNKSSSLIVKRLITRMQONHFIISVDSNQKGVGNKPKFSQKISGEGRA
VIVIPSLQVSSLEKEIKAFNNSRSHISIPFEDDKSHLHNVDDLIPVITHE
ILVOTLCWIDQADPSLHLRFLFHEYNNNSLITPKKISIFFSKENDRSLFLYNSHV
YECESVFIPLKQSHLRSTSGSFLERTHYGKIEHLVVVLGNDPXTLMLFDPDV
HYVYQKSLIASGTOFLIKKMYHLVNFQWQGYLMSQODRIHLNQLNHSYFPG
YLSVOLNNSVVRQMLENAFIMDTAIKKPFTVPIIPIIGSLAKAKKSGSGPIEK
PFRDLDSEIINFRKICKNLSHYHSGSSKKQSLYRIKPLRLSCARTLSRAKKSIV

```

```

TITLE
JOURNAL
COMMENT
113206 AATPAAAAAGTCCATTGAT 113183
Query Match 83.2%; Score 20.8; DB 15; Length 153337;
Best Local Similarity 91.7%; Pred. No. 5.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 AAAAAAAAAAGTCCATTGAT 24
113206 AATPAAAAAGTCCATTGAT 113183
RESULT 43
AC027753/c
LOCUS
DEFINITION
Homo sapiens chromosome 13 clone RP11-504H18 map 13, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
AC027753
AC027753.3 GI:8077125
VERSION
KEYWORDS
HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 155456)
Birren, B., Linton, L., Nuebaum, C. and Lander, E.
Homo sapiens chromosome 13, clone RP11-504H18
Unpublished
2 (bases 1 to 155456)
Birren, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barina, N., Bastien, V., Bada, F.,
Bogdanov, L., Bougualter, B., Brown, A., Buckler, G.,
Campopiano, A., Castele, A., Choepel, Y., Cangelosi, M., Collins, S.,
Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, D.,
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galegan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Larocque, K., Lamazeres, R., Landers, T., Lehotzky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menues, L., Mihova, T., Miranda, C., Munga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisanti, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessaye, S., Theodore, J., Tixrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7670204.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8236

```

```
Center clone name: 504_H_18
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 14436 bases at least Q40
Consensus quality: 150153 bases at least Q30
Insert size: 157000; agarose-fp
Insert size: 153156; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1055: contig of 1055 bp in length
* 1056 1155: gap of 100 bp
* 1156 2988: contig of 1833 bp in length
* 2989 3088: gap of 100 bp
* 3089 4961: contig of 1873 bp in length
* 4962 5061: gap of 100 bp
* 5062 7789: contig of 2728 bp in length
* 7790 7890: gap of 100 bp
* 7890 12206: contig of 4317 bp in length
* 12207 12306: gap of 100 bp
* 12307 16215: contig of 3909 bp in length
* 16216 16315: gap of 100 bp
* 16316 19475: contig of 3160 bp in length
* 19476 19575: gap of 100 bp
* 19576 23334: contig of 3759 bp in length
* 23335 23434: gap of 100 bp
* 23435 26922: contig of 3488 bp in length
* 26923 27022: gap of 100 bp
* 27023 30800: contig of 3778 bp in length
* 30801 30900: gap of 100 bp
* 30901 35249: contig of 4349 bp in length
* 35250 35349: gap of 100 bp
* 35350 40715: contig of 5366 bp in length
* 40716 40815: gap of 100 bp
* 40816 45789: contig of 4974 bp in length
* 45790 45889: gap of 100 bp
* 45890 51315: contig of 5426 bp in length
* 51316 51415: gap of 100 bp
* 51416 57245: contig of 5830 bp in length
* 57246 57345: gap of 100 bp
* 57346 65029: contig of 7684 bp in length
* 65030 65129: gap of 100 bp
* 65130 73057: contig of 7928 bp in length
* 73058 73157: gap of 100 bp
* 73158 81512: contig of 8355 bp in length
* 81513 81612: gap of 100 bp
* 81613 90871: contig of 9259 bp in length
* 90872 90972: gap of 100 bp
* 90973 101138: contig of 10167 bp in length
* 101139 101238: gap of 100 bp
* 101239 112252: contig of 11014 bp in length
* 112253 112352: gap of 100 bp
* 112353 122288: contig of 9936 bp in length
* 122289 122388: gap of 100 bp
* 122389 134488: contig of 12100 bp in length
* 134489 134588: gap of 100 bp
* 134589 135456: contig of 20868 bp in length.
*
* Location/Qualifiers
*     1..155456
*         /organism="Homo sapiens"
*         /mol_type="genomic DNA"
*         /db_xref="taxon:9606"
```

misc_feature /estimated_length=100
73158..81512
/note="assembly_fragment"

Query Match 83.2% Score 20.8; DB 14; Length 155456;
Match Local Similarity 91.7% Pred. No. 5.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTCAGAT 24
|||||
Db 53449 AAAAAAAAAAGTCCATTCAGAT 53426

RESULT 44
AC018647/c AC018647 160732 bp DNA linear PRI 14-OCT-2003
LOCUS Homo sapiens chromosome 7 clone RP11-379H18, complete sequence.
DEFINITION
AC018647
AC018647.5 GI:37654202
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 160732)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
Haugen,E.D.
Direct Submission
Unpublished
2 (bases 1 to 160732)
Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
Direct Submission
Submitted (15-DEC-1999) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 160732)
Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
Olson,M.V.
Direct Submission
Submitted (20-OCT-2000) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
4 (bases 1 to 160732)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
Saenphimachak,C., Buckley,D., Kibukawa,M., Raymond,C. and
Haugen,E.D.
Direct Submission
Submitted (14-OCT-2003) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Oct 14, 2003 this sequence version replaced gi:10937925.

Center: University of Washington Genome Center
Genome Center
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu

Project Information
Center project name: chr-7
Center clone name: RP11-379H18 (djs74)

Summary Statistics
Assembly program: Phrap; version 0.990319
Consensus quality: 160261 bases at least Q40
Consensus quality: 160614 bases at least Q30
Consensus quality: 160723 bases at least Q20
Insert size: 160732; sum-of-contigs
Quality coverage: 8.0x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': RP11-356G1 (UMGC:djs54) AC009333, 10610-bp overlap
3': RP11-239C9 (UMGC:djs104) AC007551, 48098-bp overlap

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.

Quality levels above 40 are expected to have less than
1 error in 10,000 bp. Base-by-base quality values are
not generally visible from the GenBank flat file format
but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

ECORI	NotI	BglII
SeqDerMap FngPrnt	SeqDerMap FngPrnt	SeqDerMap FngPrnt
-----	-----	-----
8696	9096	9736
-----	-----	-----
6	<800	813
-----	-----	-----
6603	6683	314
-----	-----	-----
5193	5218	1013
-----	-----	-----
1992	1963	9675
-----	-----	-----
10618	10861	2109
-----	-----	-----
227	<800	2151
-----	-----	-----
6313	6425	91
-----	-----	-----
1369	1392	1328
-----	-----	-----
242	<800	236
-----	-----	-----
2404	2431	3015
-----	-----	-----
2426	2431	958
-----	-----	-----
171	<800	198
-----	-----	-----
1111	1081	3367
-----	-----	-----
4972	5057	1519
-----	-----	-----
4451	4554	9215
-----	-----	-----
1603	1620	3340
-----	-----	-----
6951	7165	2446
-----	-----	-----
426	<800	516
-----	-----	-----
1453	1522	2421
-----	-----	-----
1690	1620	14980
-----	-----	-----
977	981	8773
-----	-----	-----
		8977
-----	-----	-----
		406
-----	-----	-----
		<800

4503	4554	1629	1649	4205	4269
3472	3536	3278	3290	2130	2087
5934	5922	1306	1268	1639	1637
2798	2790	5669	5660	54	<800
1538	1620	34	<800	2453	2610
741	751	3177	3290	1134	1152
2683	2683	6157	6161	817	861
3064	3082	10187	10301	1063	1063
7255	7360	4809	4862	1928	1940
2776	2790	1046	1037	1346	1341
652	<800	1708	1649	778	861
1588	1620	126	<800	14168	14370
343	<800	492	<800	1245	1236
1669	1620	6626	6540	3261	3306
458	<800	2947	2905	832	861
304	<800	4584	4626	6170	6413
18645	18764	5589	5660	6505	6413
4858	4904	12668	12989	539	<800
1409	1392	8714	8764	2591	2702
1390	1392	512	<800	8315	8547
6303	6425	23	<800	862	861
381	<800	477	<800	1521	1523
1054	1081	1459	1463	2438	2431
387	<800	5410	5482	2876	2970
952	981	394	<800	15044	15328
12333	12660	2197	2227	2331	2344
882	886			857	861
559	<800			4535	4548
1165	1160			743	<800
4811	4904				
3885	3951				
746	751				

FEATURES
source

1. 160732
/organism="Homo sapiens"
/mol_type="genomic DNA"

Query Match 83.2%; Score 20.8; DB 8; Length 160732;
Best Local Similarity 91.7%; Pred. No. 5.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 70160 AAAAAACAAGTTCATTCACAT 70137

RESULT 45
AC007839/c
LOCUS
DEFINITION
AC007839 161172 bp DNA linear INV 28-FEB-2001
Drosophila melanogaster, chromosome 2R, region 56B1-56C4, BAC clone
BACR01N09, complete sequence.
AC007839
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
1 (bases 1 to 161172)
Celinker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Bisan,D.A.,
Carlson,J.W., Center,A., Chapple,M., Chavez,C., Chew,M., Cleveland,L.,
Dodson,K., Doreelt,V., Doup,L.E., Doyle,C., Dresnek,D., Fartman,D.,
Fertler,S., Frisoe,E., Galle,R.F., Gary,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Idegian,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

TITLE
JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 161172)
Celinker,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazey,R.G.,
Butenoff,C., Champe,M., Chavez,C., Chew,M., Cleveland,L.,
Doyle,C.M., Fartman,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummel,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomoten,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Smit,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.

TITLE
JOURNAL
COMMENT
Submitted (16-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 28, 2001 this sequence version replaced gi:5670568.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

FEATURES
source

1. 161172

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/strain="Y; cn bw sp"

/db_xref="taxon:7227"

/chromosome="2R"

/map="56B1-56C4"

/clone="BACR01N09 (D656)"

/clone_lib="RP1-98 (Rosen) Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6"

ORIGIN

Query Match	83.2%	Score 20.8	DB 2	Length 161172
Best Local Similarity	91.7%	Pred. No. 5,8e+02		
Matches	22	Conservative	0	Mismatches 2
				Indels 0
				Gaps 0
QY	1	AAAAAAAAAAGTCCAAATTCAGAT	24	
Db	59738	AAAAAAAAATGTTCCAAATTAAGAT	59715	
RESULT 46				
LOCUS	EX950208	162020 bp	DNA	linear
DEFINITION	Danio rerio clone CH211-2693A6, WORKING DRAFT SEQUENCE, 9 unordered pieces.			
ACCESSION	EX950208			
VERSION	EX950208.3	GI:58578368		
KEYWORDS	HTG, HTGS, PHASE1, HTGS_DRAFT, HTGS_FULLTOP.			
SOURCE	Danio rerio (zebrafish)			
ORGANISM	Danio rerio			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.			
	1 (bases 1 to 162020)			
	McLaren,S.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (02-FEB-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk			
JOURNAL	zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk			
	On Feb 3, 2005 this sequence version replaced gi:57863412.			
COMMENT	----- Genome Center			
	Center: Wellcome Trust Sanger Institute			
	Center code: SC			
	Web site: http://www.sanger.ac.uk			
	Contact: zfish-help@sanger.ac.uk			
	----- Project Information			
	Center project name: zc2693A6			
	----- Summary Statistics			
	Assembly program: XGAP; version 4.5			
	Chemistry: Dye-terminator; 100% of reads			
	Consensus quality: 158105 bases at least Q40			
	Consensus quality: 158827 bases at least Q30			
	Consensus quality: 159248 bases at least Q20			
	Insert size: 161220; sum-of-contigs			
	Insert size: 162044; 4.2% error; agarose-fp			
	Quality coverage: 6.57% in Q20 bases; sum-of-contigs Quality coverage: 6.60% in Q20 bases; agarose-fp			

	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 9 contigs. The true order of the pieces			
	* is not known and their order in this sequence record is			
	* arbitrary. Gaps between the contigs are represented as			
	* runs of N, but the exact sizes of the gaps are unknown.			
	* This record will be updated with the finished sequence			
	* as soon as it is available and the accession number will			
	* be preserved.			

	1	2701: contig of 2701 bp in length		
	2702	2801: gap of 100 bp		
	2802	40294: contig of 37493 bp in length		
	40295	40394: gap of 100 bp		
	40395	67650: contig of 27256 bp in length		
	67651	67750: gap of 100 bp		
	67751	97838: contig of 30088 bp in length		
	97839	97938: gap of 100 bp		
	97939	102935: contig of 4397 bp in length		
	102936	103035: gap of 100 bp		
	103036	127517: contig of 24482 bp in length		
	127518	127617: gap of 100 bp		
	127618	150733: contig of 2316 bp in length		
	150734	150833: gap of 100 bp		
	150834	156681: contig of 5848 bp in length		
	156682	156781: gap of 100 bp		
	156782	162020: contig of 5239 bp in length.		
	Location/Qualifiers			

	source	1..162020	/organism="Danio rerio"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:7955"	
			/clone="CH211-269G16"	
			/clone_1fb="CHOR1-211"	
		1..2701		
	misc_feature	/note="assembly fragment:00016"		
		fragment chain:1		
		clone_end:SP6		
		vector_site:left"		
	misc_feature	2802..40294		
		/note="assembly fragment:01508"		
		fragment chain:1"		
		40395..67650		
		/note="assembly fragment:00521"		
		fragment chain:1"		
	misc_feature	67751..97838		
		/note="assembly fragment:01142"		
		fragment chain:1"		
		97939..102935		
		/note="assembly fragment:00168"		
		fragment chain:1"		
	misc_feature	103036..127517		
		/note="assembly fragment:00229"		
		fragment chain:1"		
		127618..150793		
		/note="assembly fragment:00826"		
		fragment chain:1"		
	misc_feature	150834..156681		
		/note="assembly fragment:00113"		
		fragment chain:1"		
		156782..162020		
		/note="assembly fragment:00059"		
		fragment chain:1"		
ORIGIN				
	Query Match	83.2%; Score 20.8; DB 14;	Length 162020;	
	Best Local Similarity	91.7%; Pred. No. 5.8e+02;		
	Matches 22; Conservative	0; Mismatches 2;	Indels 0;	Gaps 0;
OY	1 AAAAAAAAAAGTCCAAATTCAGAT 24			
Dd	97053 AAAAAAAAAAGTTCAATTTCAGAT 97076			
RESULT 47	AP005697/c	166423 bp	DNA	linear
LOCUS	Oryza sativa (japonica cultivar-group)			PLN 22-OCT-2003
DEFINITION	BAC clone:OSUNB0005C03.			genomic DNA, chromosome 8,
ACCESSION	AP005697			
VERSION	AP005697.3			
KEYWORDS	GI:37806333			
SOURCE	.			
ORGANISM	Oryza sativa (japonica cultivar-group)			
	Oryza sativa (japonica cultivar-group)			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatopsida; Magnoliophyta; Liliopsida; Poales; Poaceae;			
	Ehretidoideae; Oryzaceae; Oryza.			
REFERENCE	1 Sasaki,T., Matsumoto,T. and Katayose,Y.			
AUTHORS	Oryza sativa nipponbare(GM3) genomic DNA, chromosome 8, BAC			
TITLE	clone:OSUNB0005C03			
JOURNAL	Published Only in Database (2002)			
REFERENCE	2 (bases 1 to 166423)			
AUTHORS	Sasaki,T., Matsumoto,T. and Katayose,Y.			
TITLE	Direct SubMISSION			
JOURNAL	Submitted (05-SEP-2002) Takuji Sasaki, National Institute of			
	Agricobiological Sciences, Rice Genome Research Program, Kamondai			
	2-1-2, Tsukuba, Ibaraki 305-8602, Japan			
	(E-mail:tsasakia@nias.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,			
	Tel:81-298-38-7441, Fax:81-298-38-7468)			
COMMENT	On Oct 21, 2003 this sequence version replaced gi:29122869.			

Genes were predicted from the integrated results of the following:

GENSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), GENESH (<http://www.softberry.com/>), Genemark (<http://opal.biology.gatech.edu/Genemark/>), Glimmer (http://www.tigr.org/tdb/glimmer/glmr_form.html), RiceHMM (<http://rnp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor (<http://bioinformatics.lasarte.edu/cgi-bin/sp.cgi>), SIm4 (<http://www.tigr.org/software/glimmer/>), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database at RGP or DBJ. Protein homologues of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OSJNB0005C03 clone has an overlap with B1063H10 (DBJ: AP005095) clone at 5' end and with B117C04 (DBJ: AP006158) at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rnp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

```

source
    1.166423
    /organism="Oryza sativa (japonica cultivar-group)"
    /mol_type="genomic DNA"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /chromosome="8"
    /clone="OSJNB0005C03"
    join(640..799,1638..1830,1879..2374)
    /gene="OSJNB0005C03.1"
    join(640..799,1638..1830,1879..2374)
    /gene="OSJNB0005C03.1"
    /note="hypothetical ORF predicted by GENSCAN this category is not included in IRGSP standard"
gene
    complement(2781..7037)
    /gene="OSJNB0005C03.2"
    complement(2781..7037)
    /gene="OSJNB0005C03.2"
    /note="probably inactive due to including stop codon(s) in CDS"
misc_feature
    pseudogene, polypeptidein"
    /pseudo
    complement(7479..8251)
    /gene="OSJNB0005C03.3"
    complement(7479..8251)
    /gene="OSJNB0005C03.3"
    /note="probably inactive due to including frameshift(s) in CDS"
pseudogene, ORF0 protein"
    /pseudo
    complement(11503..12003)
    /gene="OSJNB0005C03.4"
    complement(11503..12003)
    /gene="OSJNB0005C03.4"
    /note="hypothetical ORF predicted by Glimmer this category is not included in IRGSP standard"
    13203..13107
gene

```

```

misc_feature
    /gene="OSJNB0005C03.5"
    13203..13107
    /gene="OSJNB0005C03.5"
    /note="probably inactive due to 3' exon missing in CDS pseudogene, retrotransposon Cnfl-1"
    /pseudo
    complement(join(15417..15587,15868..16110,16258..16692,16803..17150,17594..17980))
    /gene="OSJNB0005C03.6"
    complement(join(15417..15587,15868..16110,16258..16692,16803..17150,17594..17980))
    /gene="OSJNB0005C03.6"
    /note="retroelement-like"
    complement(19040..19369)
    /gene="OSJNB0005C03.7"
    complement(19040..19369)
    /gene="OSJNB0005C03.7"
    /note="hypothetical ORF predicted by Glimmer this category is not included in IRGSP standard"
    20793..21828
    /gene="OSJNB0005C03.8"
    join(<20793..21016,21157..21313,21364..21828)
    /gene="OSJNB0005C03.8"
    /note="start and end point are not identified"
    join(20793..21016,21157..21313,21364..21828)
    /gene="OSJNB0005C03.8"
    /note="predicted by GENSCAN etc."
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="BAC9847.1"
    /db_xref="GI:37806334"
    /translation="MCCRWNLIRVATHTATGAPVGGGRGDMQSVASCDGKWPSSGGSGMGEASVGEKTEGTGEAROVAAGAGORTGSRSAAGSAGARGRGRORRIVYRRRRRRRRGVRVRRRRRRGGSGVTHDGHGEAPATVRRRAGVARYRRAVETQYKDKGRVLRGCGNARGGNAGGGGVGVARGGARRPPRRPPPPHVAALSSAARHPAARARPLPRCRLASSTRAASTATVAVLSLPLSLVMQERERE"
    24277..25703
    /gene="OSJNB0005C03.9"
    join(<24277..24554,24974..25084,25175..25256,25332..25424,25509..25703)
    /gene="OSJNB0005C03.9"
    /note="start and end point are not identified"
    join(24277..24554,24974..25084,25175..25256,25332..25424,25509..25703)
    /gene="OSJNB0005C03.9"
    /note="similar to Oryza sativa chromosome 10, OSJNB0041L14.4"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="BAC99848.1"
    /db_xref="GI:37806335"
    /translation="MQSWSSDEVPEKYEYKPELPRLVLMCELPMPRLHMQSYIRASELGIGMTIVHPEDAFKDGPNANFAFTFKDIAFPKDKMDINLVGACLEHYILFLVYPTDQITLVLDPADYKDAVMEPLCLNLNAGRVYKRGYVXNSREKLYIRGRPCYKQPSQITNLGYYMCEMLRVNGRVTKEFDLSIPYSGQFPQKTLINLCADLCRFIRRDICNHGFEYDPHSLKLTDPKFKULREBERHAHD"
    29037..30391
    /gene="OSJNB0005C03.10"
    29037..30391
    /gene="OSJNB0005C03.10"
    /note="probably inactive due to including stop codon(s) in CDS"
pseudogene, TNP2"
    /pseudo
    32264..34028
    /gene="OSJNB0005C03.11"
    join(<32264..32317,32413..32639,32789..33168,33244..33881,33963..34028)
    /gene="OSJNB0005C03.11"
    /note="start and end point are not identified"
    join(32264..32317,32413..32639,32789..33168,33244..33881,

```

33963..34028)
/gene="OSUNJB0005C03.11"
/codon_start=1
/product="aminotransferase-like protein"
/protein_id="BAC99849.1"
/db_xref="GI:37806336"
/translation="MSSSSASPSSAAPSVAEYHEFNSNLVAIPSLSHENHYFLGPVGNPD
PTTEIIGETNRIIPRLSNPDIGHMKFTKFSKPSLEKTSPEKSKITWTFPRMDDPTLID
ITMTIGLDVNSAMPMSLINTNOYDFETKTSIGSGVAAAMGCGSVTPRHHVAFILM
WLEFLPCGSGSGPTTNQVFAETLEAKQKQPLGKILGYLYQMLNNAALAGVGV
RAAEVPRLEPIEDDERTHRCMSGGEVSTADAPAKISABILLKDWLCSFYEGFO
KDAEVPPEYDADLELPSDFRCDIMHETQTSREVTAAISCTILLPVGHQSRNI
ASVFFYHPSARQMGOLPIGLFTFDKICQCEISSTLMMDRLHLHPCPLSIEI
IELMFPSKNSDFRCWCGEMKYLFIHQPSAMWTMDFPDVVPQTSTSSPPRGNSGRDIT
YAQS"
34035..35400
/gene="OSUNJB0005C03.12"
join(<34033..34127,34186..34299,34456..34515,34775..34972
35057..35268,35361..35400)
/gene="OSUNJB0005C03.12"
/note="start and end point are not identified"
join(34035..34127,34186..34299,34456..34515,34775..34972
35057..35268,35361..35400)
/gene="OSUNJB0005C03.12"
/codon_start=1
/product="aminotransferase-like"
/protein_id="BAC99850.1"
/db_xref="GI:37806337"
/translation="MEVGYLLLSAMMPRLQLFPRKDSSENGPMPKIKKAKPADDLPL

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
3	Research, 320 Charles Street, Cambridge, MA 02141, USA			
3	(bases 1 to 170030)			
	Bliren,B., Lincoln,U., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Colliamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gaidyna,S., Geraiidery,K., Grant,G., Hages,B., Heaford,A., Herrens,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwen,P., Mcgurk,A., McKernan,K., Meldim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J., Nahf,R., Naylor,J., Nioff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stillewell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Tortorella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,U. and Zody,M.			
	Direct Submission			
	Submitted (15-NOV-1998)			
	Research, 320 Charles Street, Cambridge, MA 02141, USA			
	On Nov 15, 1998 this sequence version replaced g1:368742.			
	All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)			
	http://ftp.genome.washington.edu/RM/RepeatMasker.html			
	Only the first 170030 base pairs of this clone are being submitted. The remainder overlaps accession number AC004139 (WICGR project L291).			
FEATURES				
SOURCE				
	Location/Qualifiers			
	1..170030			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
	/chromosome="17"			
	/map="17"			
	/clone="hRPK_259_G_18"			
	/clone_lib="RPCT-11 human BAC library"			
	2..271			
	/rpt_family="L1ME"			
	476..786			
	/rpt_family="Aluub"			
	1750..1915			
	/rpt_family="T2"			
	complement(2143..2448)			
	/rpt_family="AluY"			
	complement(2600..2810)			
	/rpt_family="MIR"			
	complement(5987..6018)			
	/rpt_family="(CCCTA)n"			
	6704..6987			
	/rpt_family="AluSg"			
	6988..7007			
	/rpt_family="(TAAA)n"			
	7141..7173			
	/rpt_family="AT rich"			
	complement(7360..7633)			
	/rpt_family="Tiggrs(Golem)"			
	7634..7942			
	/rpt_family="AluSx"			
	complement(7943..8051)			
	/rpt_family="Tiggrs(Golem)"			
	complement(8528..8838)			
	/rpt_family="AluSx"			
	complement(8901..9240)			
	/rpt_family="Aluub"			
	10142..10422			
	/rpt_family="AluSg"			
	10473..10749			
	/rpt_family="Aluub"			
	complement(10750..10769)			
	/rpt_family="(CAAA)n"			
	complement(10771..11044)			
	/rpt_family="Aluub"			

```

repeat_region 11249..11366
/rpt_family="AluSg/x"
repeat_region 1137..11391
/rpt_family="AT_rich"
repeat_region 11417..11703
/rpt_family="Aluub"
repeat_region 11888..12188
/rpt_family="AluSg"
repeat_region 12189..12263
/rpt_family="MER41B"
repeat_region 12275..12290
/rpt_family="Aluub"
repeat_region 12291..12582
/rpt_family="AluSg"
repeat_region 12583..12731
/rpt_family="Aluub"
repeat_region 15348..15659
/rpt_family="AluYb8"
repeat_region 15721..16004
/rpt_family="AluJo"
repeat_region 16307..17274
/rpt_family="LTR5"
repeat_region 17651..17753
/rpt_family="LIMC/D"
repeat_region complement(17754..18051)
/rpt_family="AluSc"
repeat_region 18052..18063
/rpt_family="LIMC/D"
repeat_region 18167..18481
/rpt_family="AluSx"
repeat_region 18486..18796
/rpt_family="AluY"
repeat_region 18846..18873
/rpt_family="AT_rich"
repeat_region 18945..19239
/rpt_family="AluSx"
repeat_region 19243..19509
/rpt_family="AluJo"
repeat_region complement(19558..19700)
/rpt_family="L2"
repeat_region 19709..19968
/rpt_family="HAL1"
repeat_region 20013..20318
/rpt_family="AluSg"
repeat_region 20343..20369
/rpt_family="AT_rich"
repeat_region complement(20370..20495)
/rpt_family="FLAM_C"
repeat_region complement(20502..20797)
/rpt_family="AluSg"
repeat_region 20798..20815
/rpt_family="AT_rich"
repeat_region 20866..21093
/rpt_family="HAL1"
repeat_region 21094..21398
/rpt_family="AluSg"
repeat_region 21405..21715
/rpt_family="Aluub"
repeat_region 21716..21785
/rpt_family="AluSx"
repeat_region 21791..22093
/rpt_family="HAL1"
repeat_region 22187..22465
/rpt_family="AluSg"
repeat_region 22466..22521
/rpt_family="(CAAA)n"
repeat_region 22643..22938
/rpt_family="AluSx"
repeat_region 22947..23251
/rpt_family="AluSg"
repeat_region 23784..24097
/rpt_family="Aluub"
repeat_region 24698..24993

```

```

repeat_region /rpt_family="AluSg"
25122..25147
/rpt_family="(CA)n"
repeat_region /rpt_family="AluSg"
25698..26013
repeat_region /rpt_family="AluSg"
26078..26104
repeat_region /rpt_family="AT_rich"
complement(27061..27091)
/rpt_family="(CAAA)n"
repeat_region 28173..28481
/rpt_family="AluSg"
repeat_region 28829..29117
/rpt_family="AluSg"
repeat_region 29407..29727
/rpt_family="AluSg"
repeat_region complement(29974..30261)
/rpt_family="AluSg"
repeat_region 31297..31320
/rpt_family="AT_rich"
repeat_region complement(31353..31636)
/rpt_family="AluSg"
repeat_region 31885..31907

```

Query Match 83.2%; Score 20.8; DB 8; Length 170030;
 Best Local Similarity 91.7%; Pred. No. 5.7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 2 AAAAAAAAAAGTTCATTTCAGATA 25
Db 131487 AAAAAAAAAAGTTCATTTCAGATA 131464

```

```

RESULT 49
AC024604 172546 bp DNA linear HTG 30-AUG-2001
LOCUS Homo sapiens chromosome 10 clone RP11-331G19, WORKING
DEFINITION DRAFT SEQUENCE, 22 unordered pieces.
ACCESSION AC024604 GI:8389428
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 172546)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172546)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Jun 9, 2000 this sequence version replaced gi:7549605.
-----
Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
-----
Project Information
Center project name: hg202
-----
Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 152033 bases at least Q40
Consensus quality: 163535 bases at least Q30
Consensus quality: 165256 bases at least Q20
Insert size: 170446; sum-of-coverage
Quality coverage: 4.3x in Q20 bases; sum-of-coverage

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1339: contig of 1339 bp in length
* 1340 1439: gap of unknown length
* 1440 2789: contig of 1350 bp in length
* 2790 2889: gap of unknown length
* 2890 4562: contig of 1673 bp in length
* 4563 4662: gap of unknown length
* 4663 6444: contig of 1682 bp in length
* 6445 9221: contig of 2777 bp in length
* 9222 9321: gap of unknown length
* 9322 11997: contig of 2676 bp in length
* 11998 12097: gap of unknown length
* 12098 14867: contig of 2770 bp in length
* 14868 14967: gap of unknown length
* 14968 19919: contig of 4952 bp in length
* 19920 20019: gap of unknown length
* 20020 24125: contig of 4106 bp in length
* 24126 24225: gap of unknown length
* 24226 29444: contig of 5219 bp in length
* 29445 34568: gap of unknown length
* 34569 34668: contig of 5024 bp in length
* 34669 39414: gap of unknown length
* 39415 39514: gap of unknown length
* 39515 45885: contig of 6371 bp in length
* 45886 45985: gap of unknown length
* 45986 50997: contig of 5012 bp in length
* 50998 51097: gap of unknown length
* 51098 59143: contig of 8046 bp in length
* 59144 59243: gap of unknown length
* 59244 67779: contig of 8536 bp in length
* 67780 67880: gap of unknown length
* 67881 76429: contig of 8550 bp in length
* 76430 76530: gap of unknown length
* 76531 86704: contig of 10175 bp in length
* 86705 86804: gap of unknown length
* 86805 103147: contig of 16343 bp in length
* 103148 103247: gap of unknown length
* 103248 116343: contig of 13096 bp in length
* 116344 116443: gap of unknown length
* 116444 135318: contig of 18875 bp in length
* 135319 135418: gap of unknown length
* 135419 172546: contig of 37128 bp in length.
*
Location/Qualifiers
1. 172546
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="Chromosome 10"
/clone="RP11-331G19"
/clone_1fb="RP11-331G19"
/clone_2fb="RP11-331G19"
1340_1439
/estimated_length=unknown
2790_2889
/estimated_length=unknown
4563_4662
/estimated_length=unknown
6345_6444
/estimated_length=unknown
9222_9321
/estimated_length=unknown
11998_12097
/estimated_length=unknown
14868_14967
/estimated_length=unknown
```

```
gap 19920..20019
/estimated_length=unknown
gap 24126..24225
/estimated_length=unknown
gap 29445..29544
/estimated_length=unknown
gap 34569..34668
/estimated_length=unknown
gap 39415..39514
/estimated_length=unknown
gap 45886..45985
/estimated_length=unknown
gap 50998..51097
/estimated_length=unknown
gap 59144..59243
/estimated_length=unknown
gap 67780..67879
/estimated_length=unknown
gap 76430..76529
/estimated_length=unknown
gap 86705..86804
/estimated_length=unknown
gap 103148..103247
/estimated_length=unknown
gap 116344..116443
/estimated_length=unknown
gap 135319..135418
/estimated_length=unknown
```

ORIGIN

```
Query Match 83.2%; Score 20.8; DB 14; Length 172546;
Best Local Similarity 91.7%; Pred. No. 5.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 AAAAAAAAAAGTCCATTGACAT 24

Db 138387 AAAAAAAAAAGTCCATTGACAT 138410

RESULT 50

AC166110

LOCUS AC166110 174420 bp DNA linear HTG 23-JUL-2005
DEFINITION Mus musculus chromosome 17 clone RP23-130A4, *** SEQUENCING IN
PROGRESS ***; 57 unordered pieces.

ACCESSION AC166110.1 GI:71067315

VERSION HTG; HTGS PHASE1.

KEYWORDS Mus musculus (house mouse)

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
----- Project Information -----
Center project name: W_BA0130A04
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-Primer ET; 0% of reads
```

Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 140397 bases at least Q40
Consensus quality: 148231 bases at least Q30
Consensus quality: 154542 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1527: contig of 1527 bp in length
1528 1627: gap of unknown length
1628 3163: contig of 1536 bp in length
3164 3263: gap of unknown length
3264 4425: contig of 1162 bp in length
4426 4525: gap of unknown length
4526 5805: contig of 1280 bp in length
5806 5905: gap of unknown length
5906 7181: contig of 1276 bp in length
7182 7282: gap of unknown length
7282 8683: contig of 1402 bp in length
8684 8783: gap of unknown length
8784 10684: contig of 1901 bp in length
10685 10784: gap of unknown length
10785 12417: contig of 1633 bp in length
12418 12517: gap of unknown length
12517 13807: contig of 1290 bp in length
13808 13907: gap of unknown length
13908 15311: contig of 1404 bp in length
15312 15411: gap of unknown length
15412 16780: contig of 1369 bp in length
16781 16880: gap of unknown length
16881 17980: contig of 1100 bp in length
17981 18080: gap of unknown length
18081 19639: contig of 1559 bp in length
19640 19739: gap of unknown length
19740 21650: contig of 1911 bp in length
21651 21751: gap of unknown length
21751 23936: contig of 2186 bp in length
23937 24036: gap of unknown length
24037 26321: contig of 2285 bp in length
26322 28349: contig of 1928 bp in length
28350 28449: gap of unknown length
28450 30629: contig of 2180 bp in length
30630 30729: gap of unknown length
30730 32506: contig of 1777 bp in length
32507 32606: gap of unknown length
32607 34412: contig of 1806 bp in length
34413 34512: gap of unknown length
34513 35777: contig of 1265 bp in length
35778 35877: gap of unknown length
35879 38175: contig of 2298 bp in length
38176 38275: gap of unknown length
38276 40446: contig of 2171 bp in length
40447 40546: gap of unknown length
40547 43413: contig of 2867 bp in length
43414 43513: gap of unknown length
43514 44971: contig of 1458 bp in length
44972 45071: gap of unknown length
45072 47799: contig of 2728 bp in length
47800 47899: gap of unknown length
47900 49561: contig of 1662 bp in length
49562 49661: gap of unknown length
49662 51344: contig of 1683 bp in length
51345 51444: gap of unknown length
51445 55158: contig of 3714 bp in length
55159 55258: gap of unknown length
55260 57823: contig of 2565 bp in length

57824 57923: gap of unknown length
57924 60912: contig of 2989 bp in length
60913 61012: gap of unknown length
61013 62864: contig of 1852 bp in length
62865 62964: gap of unknown length
62965 66651: contig of 3687 bp in length
66652 66751: gap of unknown length
66752 70549: contig of 3798 bp in length
70550 71112: contig of 3463 bp in length
71113 74213: gap of unknown length
74213 76427: contig of 2215 bp in length
76428 76527: gap of unknown length
76528 79348: contig of 2821 bp in length
79349 79448: gap of unknown length
79449 81017: contig of 1569 bp in length
81018 81117: gap of unknown length
81118 84006: contig of 2889 bp in length
84007 84106: gap of unknown length
84107 87000: contig of 2894 bp in length
87001 89684: contig of 2584 bp in length
89685 89784: gap of unknown length
89785 91234: contig of 1450 bp in length
91235 91335: gap of unknown length
91335 94098: contig of 2764 bp in length
94099 94198: gap of unknown length
94199 97421: contig of 3223 bp in length
97422 97521: gap of unknown length
97522 101829: contig of 4308 bp in length
101830 101929: gap of unknown length
101930 104956: contig of 3027 bp in length
104957 105056: gap of unknown length
105057 109171: contig of 4115 bp in length
109172 109271: gap of unknown length
109272 116382: contig of 7111 bp in length
116383 116482: gap of unknown length
116483 120918: contig of 4436 bp in length
120919 124018: gap of unknown length
124019 124723: contig of 3705 bp in length
124724 124823: gap of unknown length
124824 131472: contig of 6649 bp in length
131473 131572: gap of unknown length
131573 138395: contig of 6823 bp in length
138396 138495: gap of unknown length
138496 145033: contig of 6538 bp in length
145034 145133: gap of unknown length
145134 151548: contig of 6415 bp in length
151549 151648: gap of unknown length
151649 157688: contig of 6040 bp in length
157689 157788: gap of unknown length
157789 166243: contig of 8455 bp in length
166244 166344: gap of unknown length
166345 174420: contig of 8077 bp in length.
Location/Qualifiers
1. 174420
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-130A4"
1. 1527
/note="assembly_name:Contig17"
1528. 1627
/estimated_length=unknown
1628. 3163
/note="assembly_name:Contig19"
3164. 3263
/estimated_length=unknown
3264. 4425
/note="assembly_name:Contig20"
4426. 4525
/estimated_length=unknown

misc_feature 4526..5805
/note="assembly_name:Contig24"
gap 5806..5905
/estimated_length=unknown
misc_feature 5906..7181
/note="assembly_name:Contig34"
gap 7182..7281
/estimated_length=unknown
misc_feature 7282..8683
/note="assembly_name:Contig35"
gap 8684..8783
/estimated_length=unknown
misc_feature 8784..10684
/note="assembly_name:Contig46"
gap 10685..10784
/estimated_length=unknown
misc_feature 10785..12417
/note="assembly_name:Contig47"
gap 12418..12517
/estimated_length=unknown
misc_feature 12518..13807
/note="assembly_name:Contig48"

Query Match 83.2%; Score 20.8; DB 14; Length 174420;
Best Local Similarity 91.7%; Pred. No. 5.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCAAATTCAGAT 24
Db 58870 AAAAAAAAAAGTCCAAATTCAGAT 58893

Search completed: December 14, 2005, 11:11:45
Job time : 874.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 ; Search time 203.2 Seconds
(without alignments) 819.967 Million cell updates/sec

Title: US-10-681-773-8

Perfect score: 25
Sequence: 1 aaaaaaaaaagttccaatcagata 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N_Geneseq_21:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.8	83.2	2567	13	ADXS5035
2	20.8	83.2	31271	4	ABL26680
3	20.2	80.8	3759	4	ABL08968
4	20.2	80.8	3759	10	ADCS35776
5	20.2	80.8	3759	10	ADCS35776
6	20.2	80.8	4169	4	ABL16690
7	20.2	80.8	5933	6	ABK39977
8	20.2	80.8	5933	6	ABK39977
9	20.2	80.8	6486	6	ABQ67050
10	20.2	80.8	9510	6	ABQ67050
11	20.2	80.8	9510	6	ABQ67050
12	20.2	80.8	9510	6	ABQ67050
13	20.2	80.8	10151	6	ABL34432
14	20.2	80.8	10151	6	ABL34432
15	20.2	80.8	12542	8	ABX74149
16	20.2	80.8	12542	8	ABX74149
17	20.2	80.8	15332	4	AAK78660
18	20.2	80.8	15332	4	AAK78660
19	20.2	80.8	20142	8	ACA19757

C	20	20.2	80.8	109586	11	ACN43994	Acn43994 Human gen
C	21	19.8	79.2	7491	3	AAA70156	Aaa70156 Plasmodiu
C	22	19.4	77.6	665	4	AAL25065	Aal25065 Human bre
C	23	19.2	76.8	42	3	AACS5139	Aacs5139 Arabidops
C	24	19.2	76.8	424	4	AAK72039	Aak72039 Human imm
C	25	19.2	76.8	533	5	ABV10721	Abv10721 Human pro
C	26	19.2	76.8	724	10	ADB61077	Adb61077 Human gen
C	27	19.2	76.8	724	10	ADB60529	Adb60529 Human gen
C	28	19.2	76.8	832	5	ABV11370	Abv11370 Human pro
C	29	19.2	76.8	1356	3	AACS9496	Aacs9496 Human sec
C	30	19.2	76.8	1654	11	ACU34950	Act34950 Rice stre
C	31	19.2	76.8	2000	11	ACU36184	Act36184 Rice stre
C	32	19.2	76.8	3012	4	ABL28464	AbL28464 Drosophi1
C	33	19.2	76.8	3453	6	ABA05542	AbA05542 Murine a1
C	34	19.2	76.8	6392	6	ABL32684	AbL32684 Human imm
C	35	19.2	76.8	6392	6	ABL34506	AbL34506 Human met
C	36	19.2	76.8	6392	7	ADS99767	AdS99767 BieluPhic
C	37	19.2	76.8	9707	6	ABL33421	AbL33421 Human imm
C	38	19.2	76.8	10695	4	AAK5420	Aak5420 Human imm
C	39	19.2	76.8	18664	4	AAK84438	Aak84438 Human imm
C	40	19.2	76.8	18664	4	AAK5421	Aak5421 Human imm
C	41	19.2	76.8	32231	9	ABK42519	Abk42519 Genomic s
C	42	19.2	76.8	32231	9	ADB60675	AdB60675 Connectiv
C	43	19.2	76.8	66788	13	ADV15296	Adv15296 Human otc
C	44	19.2	76.8	90120	12	ADQ97413	Adq97413 Human can
C	45	19.2	76.8	115829	13	ABD33448	Abd33448 Human can
C	46	18.8	75.2	3696	4	AAC90666	Aac90666 Streptococ
C	47	18.8	75.2	235962	14	ADZ12926	Adz12926 Murine ca
C	48	18.6	74.4	102	2	AAV75702	Aav75702 Steaphyloc
C	49	18.6	74.4	121	12	ADK94078	Adk94078 Polynucle
C	50	18.6	74.4	123	8	ABX79682	Abx79682 EST polym
C	51	18.6	74.4	265	2	AAV75430	Aav75430 Steaphyloc
C	52	18.6	74.4	300	2	AAZ14466	Aaz14466 Human gen
C	53	18.6	74.4	348	4	AAI89378	Aai89378 Human pol
C	54	18.6	74.4	417	5	ABV23480	Abv23480 Human pro
C	55	18.6	74.4	417	5	ABV29338	Abv29338 Human pro
C	56	18.6	74.4	422	5	ABV14672	Abv14672 Human pro
C	57	18.6	74.4	424	14	ACL63280	ACL63280 Human col
C	58	18.6	74.4	451	5	ABV05503	Abv05503 Human pro
C	59	18.6	74.4	463	5	ABV44553	Abv44553 Human pro
C	60	18.6	74.4	463	5	ABV35750	Abv35750 Human pro
C	61	18.6	74.4	500	3	AAC93597	Aac93597 Cat flea
C	62	18.6	74.4	536	13	ADQ57909	Adq57909 Novel can
C	63	18.6	74.4	798	2	AAZ15515	Aaz15515 Human gen
C	64	18.6	74.4	945	1	AAI20961	Aai20961 Human bre
C	65	18.6	74.4	951	6	ABN66936	Abn66936 Streptococ
C	66	18.6	74.4	1315	7	ADZ74629	Adz74629 Arabidops
C	67	18.6	74.4	1386	13	ADT16012	Adt16012 Plant cdn
C	68	18.6	74.4	1898	12	ADQ23073	Adq23073 Human sof
C	69	18.6	74.4	2020	10	ADB63017	AdB63017 Human cdn
C	70	18.6	74.4	2044	6	ABQ67008	Abq67008 Human ang
C	71	18.6	74.4	2663	4	ABL25550	AbL25550 Drosophi1
C	72	18.6	74.4	3097	11	ADM02521	Adm02521 Human cdn
C	73	18.6	74.4	3097	13	ABK64551	Abk64551 Human lup
C	74	18.6	74.4	3134	4	ABL27132	AbL27132 Drosophi1
C	75	18.6	74.4	3628	10	ADP55406	AdP55406 Human DNA
C	76	18.6	74.4	5338	6	ABK31378	Abk31378 Signal tr
C	77	18.6	74.4	5338	6	ABL70329	AbL70329 Chemical1
C	78	18.6	74.4	5338	6	AA661280	Aa661280 Human gen
C	79	18.6	74.4	5557	6	ABU33546	AbU33546 Human imm
C	80	18.6	74.4	5754	6	ABN33521	AbN33521 Human imm
C	81	18.6	74.4	5773	6	ABN80261	Abn80261 Human che
C	82	18.6	74.4	5982	6	ABJ34049	AbJ34049 Human imm
C	83	18.6	74.4	6012	6	ABK28338	Abk28338 DNA trans
C	84	18.6	74.4	6015	6	ABU32676	AbU32676 Human imm
C	85	18.6	74.4	6045	2	AAU74665	Aau74665 Ureophan
C	86	18.6	74.4	6059	4	AAU74665	Aau74665 Nucleotid
C	87	18.6	74.4	6283	6	ABK31434	AbK31434 Signal tr
C	88	18.6	74.4	6283	6	ABL170403	AbL170403 Chemical1
C	89	18.6	74.4	6283	6	AA661351	Aa661351 Human gen
C	90	18.6	74.4	6316	6	ABL33639	AbL33639 Human imm
C	91	18.6	74.4	6447	6	ABL33087	AbL33087 Human imm
C	92	18.6	74.4	6447	6	ABL34533	AbL34533 Human met

```
C 93 18.6 74.4 6447 6 ABL70548 ABL70548 Chemicali
C 94 18.6 74.4 6447 6 ABQ66974 ABQ66974 Human ang
C 95 18.6 74.4 6447 7 ADS99794 ADS99794 Complemen
C 96 18.6 74.4 6721 6 ABL33529 ABL33529 Human imm
C 97 18.6 74.4 7001 13 ADS89476 ADS89476 Oligonuc
C 98 18.6 74.4 7113 6 ABL32805 ABL32805 Human imm
C 99 18.6 74.4 7131 6 ABL31450 ABL31450 Signal tr
C 100 18.6 74.4 7131 6 ABL70427 ABL70427 Chemicali
C 101 18.6 74.4 7131 6 ABL31360 ABL31360 Human gen
C 102 18.6 74.4 7200 8 ACA46122 ACA46122 Prokaryot
C 103 18.6 74.4 8002 6 AAD31629 AAD31629 Arabidops
C 104 18.6 74.4 8129 4 AAS46763 AAS46763 Tumour su
C 105 18.6 74.4 8781 6 ABL33687 ABL33687 Human imm
C 106 18.6 74.4 8964 6 ABL33535 ABL33535 Human imm
C 107 18.6 74.4 8964 6 ABL31373 ABL31373 Signal tr
C 108 18.6 74.4 8964 6 ABL70582 ABL70582 Chemicali
C 109 18.6 74.4 8964 6 AAS61277 AAS61277 Human gen
C 110 18.6 74.4 9353 8 ABZ68800 ABZ68800 Nucleotid
C 111 18.6 74.4 9524 8 ABL33993 ABL33993 Human che
C 112 18.6 74.4 9524 6 ABL32838 ABL32838 Human che
C 113 18.6 74.4 9733 6 ABL32683 ABL32683 Human imm
C 114 18.6 74.4 10254 6 ABQ67046 ABQ67046 Human ang
C 115 18.6 74.4 10302 6 ABK81961 ABK81961 CDNA enco
C 116 18.6 74.4 10302 6 ABS68901 ABS68901 Human utr
C 117 18.6 74.4 10302 13 ADPS5038 ADPS5038 Human PRO
C 118 18.6 74.4 10320 2 AAT74666 AAT74666 Full-leng
C 119 18.6 74.4 10386 5 AAS76114 AAS76114 DNA encod
C 120 18.6 74.4 10597 8 ABZ68799 ABZ68799 Nucleotid
C 121 18.6 74.4 10853 8 ABZ68805 ABZ68805 Nucleotid
C 122 18.6 74.4 11398 6 ABL33439 ABL33439 Human imm
C 123 18.6 74.4 11422 6 ABL33936 ABL33936 Human che
C 124 18.6 74.4 11422 6 ABL3218 ABL3218 Human imm
C 125 18.6 74.4 13024 13 ADPS5623 ADPS5623 Human PRO
C 126 18.6 74.4 13558 4 AAI66274 AAI66274 Murine ap
C 127 18.6 74.4 14861 6 ABL33194 ABL33194 Human imm
C 128 18.6 74.4 14861 6 ABL70553 ABL70553 Chemicali
C 129 18.6 74.4 14861 6 AAS61202 AAS61202 Human gen
C 130 18.6 74.4 15686 6 AAD38801 AAD38801 Rice RGAA
C 131 18.6 74.4 16033 6 ABL33404 ABL33404 Human imm
C 132 18.6 74.4 16532 6 ABZ35556 ABZ35556 Human gen
C 133 18.6 74.4 17893 6 ABL33364 ABL33364 Human imm
C 134 18.6 74.4 18011 6 ABL32035 ABL32035 Human imm
C 135 18.6 74.4 18846 4 AAS55261 AAS55261 Staphyloc
C 136 18.6 74.4 19205 5 AAS34685 AAS34685 Human DNA
C 137 18.6 74.4 24477 14 AEB39174_09 AEB39174_09
C 138 18.6 74.4 28605 2 ACF74568 ACF74568 Staphyloc
C 139 18.6 74.4 31096 2 AAV74370 AAV74370 Staphyloc
C 140 18.6 74.4 31497 8 ABL15082 ABL15082 Pathogen
C 141 18.6 74.4 32167 4 AAK90553 AAK90553 Human dig
C 142 18.6 74.4 32169 5 ABA14358 ABA14358 Human ner
C 143 18.6 74.4 32195 5 AAS32232 AAS32232 Human DNA
C 144 18.6 74.4 33195 6 ABB67535 ABB67535 Novel hum
C 145 18.6 74.4 35641 6 ABL64428 ABL64428 Stomach c
C 146 18.6 74.4 35641 6 ABB95727 ABB95727 Gene #222
C 147 18.6 74.4 39119 8 ABZ74034 ABZ74034 Secretd
C 148 18.6 74.4 39119 8 ADA98641 ADA98641 Human sec
C 149 18.6 74.4 39119 10 ADC20764 ADC20764 Human sec
C 150 18.6 74.4 39119 10 ABZ67621 ABZ67621 Human sec
```

ALIGNMENTS

```
RESULT 1
ID ADX55035/c
ADX55035 standard; cDNA; 2567 BP.
```

```
XX ADX55035;
```

```
XX 21-APR-2005 (first entry)
```

```
XX Plant full length insert polynucleotide seqid 29775.
```

```
KM plant protectant; plant growth regulant; gene therapy; plant;
KM recombinant DNA construct; physical array; plant breeding marker;
KM cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KM extreme osmotic condition; pathogen tolerance; pest tolerance;
KM growth rate; cell cycle pathway; disease resistance;
KM galactomannan production; lignin production; plant growth regulator;
KM yield; plant growth; plant development; seed oil; protein yield;
KM protein content; gene; ss.
```

Unidentified.

US2004034888-A1.

19-FEB-2004.

28-APR-2003; 2003US-00425114.

06-MAY-1999; 99US-00304517.

05-NOV-2001; 2001US-00985678.

(LIU/) LIU J.

(KHOV/) ZHOU Y.

(SCRE/) SCREEN S E.

(TABAS/) TABASKA J E.

(CAOY/) CAO Y.

LIU J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

WPI, 2004-180133/17.

New recombinant DNA construct, useful for improving plant tolerance to

cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or

pests, for conferring increased resistance to plant disease, or for

improving yield.

Claim 1; SEQ ID NO 29775; 15pp; English.

The invention describes a recombinant DNA construct comprising a

polynucleotide consisting of a sequence encoding an amino acid sequence

available in electronic form from the US patent office at

ftp.segdata.uspto.gov/sequence.html?docid=2004034888. The polynucleotide

of the invention are also useful in physical arrays of molecules and as

plant breeding markers. The recombinant DNA construct is useful for

improving plant tolerance to cold, heat, drought, herbicides, extreme

osmotic conditions, pathogens or pests, for manipulating growth rate in

plant cells by modification of the cell cycle pathway, for conferring

increased resistance to plant disease, for producing galactomannan,

lignin or plant growth regulators, for increasing the rate of homologous

recombination in plants, for improving yield by modification of

photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake

or by providing improved plant growth and development under at least one

stress condition or for modifying seed oil or protein yield and/or

content. This sequence represents a plant full length insert

polynucleotide that can be used in the recombinant DNA construct of the

invention.

Sequence 2567 BP; 699 A; 461 C; 524 G; 883 T; 0 U; 0 Other;

Query Match 83.2%; Score 20.8; DB 13; Length 2567;

Best Local Similarity 91.7%; Pred. No. 2e+02; Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 AAAAAAAAAAGTTCATTCAGATA 25

2388 AAAAAAAAAAGTTCATTCAGATA 2365

RESULT 2

ABL26680

ABL26680;


```

XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31513.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
PA (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 31513; 21pp + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 31271 BP; 8794 A; 6300 C; 6590 G; 9587 T; 0 U; 0 Other;
SQ
Query Match 83.2%; Score 20.8; DB 4; Length 31271;
Best Local Similarity 91.7%; Pred. No. 2.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTTCATTTCAGAT 24
Db 6323 AAAAAAAAAAGTTCATTTCAGAT 6346

```

```

XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR P-PSDB; ABB64865.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 21386; 21pp + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3759 BP; 1081 A; 715 C; 763 G; 1200 T; 0 U; 0 Other;
SQ
Query Match 80.8%; Score 20.2; DB 4; Length 3759;
Best Local Similarity 88.0%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTTCATTTCAGATA 25
Db 303 AAAAAAAAAAGTTCATTTCAGATA 279

```

```

RESULT 3
ABU08968/C
ID ABL08968 standard; cDNA; 3759 BP.
AC
XX ABL08968;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 21386.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF

```

```

RESULT 4
AAS57090/C
ID AAS57090 standard; DNA; 3759 BP.
AC
XX AAS57090;
XX
XX 16-JAN-2002 (first entry)
DT
XX DNA encoding Drosophila G-protein coupled receptor, GCPR #10.
DE
XX Drosophila; G-protein coupled receptor; GCPR; insecticide; diagnostic;
XX mutation detection; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200170980-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009341.
PF
XX 23-MAR-2000; 2000US-0191638P.
PR 18-JUL-2000; 2000US-00618893.
XX
XX (PEKE ) PE CORP NY.
XX
XX Cravchik A;
XX WPI; 2001-616405/71.
DR P-PSDB; AAU38932.
XX
XX Sixty six Drosophila Melanogaster G-protein coupled receptors (GPCR),
PT useful in the treatment and diagnosis of GPCR-related conditions and for
PT identifying GPCR modulators for use as insecticides.

```

XX Claim 4; Page 85-86; 392pp; English.
PS
XX
CC The invention relates to sixty six novel isolated *Drosophila melanogaster*
CC G-protein coupled receptors (GPCR). The GPCR proteins and nucleic acids
CC are useful in the treatment and diagnosis of GPCR-related conditions. The
CC GPCR proteins and nucleic acids are also useful for identifying
CC modulators of GPCR proteins for use as insecticides. The nucleic acid can
CC also be used to detect mutations in GPCR genes and gene expression
CC products such as mRNA. AAS57072-AAS57203 represent *D. melanogaster* G-
CC coupled protein receptor genomic and coding sequences of the invention
XX
SQ Sequence 3759 BP; 1081 A; 715 C; 763 G; 1200 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 4; Length 3759;
Best Local Similarity 88.0%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 303 AAAAAAAAAACTTACATTTCAGATA 279
RESULT 5
ADC35776/C
ID ADC35776 standard; DNA; 3759 BP.
XX
AC ADC35776;
XX
DT 18-DEC-2003 (first entry)
XX
DE *Drosophila* G protein coupled receptor genomic DNA seq id 10.
XX
KW G-protein coupled receptor; GPCR; insecticide; drug screening;
KW insecticide screening; insecticidal activity; insecticidal tolerance;
KW fruit fly; gene; ds.
XX
OS *Drosophila melanogaster*.
XX
PN US2003092124-A1.
XX
PD 15-MAY-2003.
XX
PF 15-OCT-2002; 2002US-00270333.
XX
PR 03-DEC-1999; 99US-0168677P.
XX
PR 12-JAN-2000; 2000US-0175691P.
XX
PR 23-MAR-2000; 2000US-0191638P.
XX
PR 18-JUL-2000; 2000US-00618893.
XX
PA (APPL-) APPLERA CORP.
XX
PI Cravchik A;
XX
DR WPI: 2003-765480/72.
XX
DR P-85DB; ADC35778.
XX
PT New isolated G-protein coupled receptor useful for identifying modulators
XX protein and for identifying compounds that modulate receptor activity.
XX
PS Claim 4; SEQ ID NO 28; 130pp; English.
XX
XX The invention describes an isolated protein (I) consisting or comprising
CC an amino acid sequence selected from fully defined 66 G-protein coupled
CC receptor amino acid sequences (SI), as given in the specification, an
CC allelic variant of (SI), an orthologue of (SI) or fragment of (SI). (I)
CC is useful for identifying an agent that binds to (I) which comprises
CC contacting the protein with an agent and assaying the contacted mixture
CC to determine whether a complex is formed with the agent bound to the
CC protein. (I) is useful for identifying modulators as potential
CC insecticides, to determine the biological activity of the protein (a
CC panel of multiple proteins for high-throughput screening), as targets for

CC identifying agents for use in human drugs and for identifying compounds
CC that modulate receptor activity. An antibody (II) that selectively binds
CC to (I) is useful for assessing normal and aberrant subcellular
CC localisation of cells and monitoring a treatment modality. A nucleic acid
CC (III) encoding (I) is useful for drug/insecticide screening to identify
CC compounds that modulate G-protein coupled receptor (GPCR) nucleic acid
CC expression, diagnostic assays for qualitative changes in GPCR nucleic
CC acid that lead to insecticidal activity/tolerance, to detect mutations in
CC GPCR genes and gene expression products such as mRNA, and as
CC hybridisation probes for determining the presence, level, form and
CC distribution of nucleic acid expression. A host cell comprising a vector
CC containing (III) is useful for conducting cell-based assays involving the
CC GPCR protein or its fragments, and identifying GPCR protein mutants. This
CC sequence encodes a fruit fly G-protein coupled receptor (GPCR).
XX
SQ Sequence 3759 BP; 1081 A; 715 C; 763 G; 1200 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 10; Length 3759;
Best Local Similarity 88.0%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 303 AAAAAAAAAACTTACATTTCAGATA 279
RESULT 6
ABL16690/C
ID ABL16690 standard; DNA; 4169 BP.
XX
AC ABL16690;
XX
DT 26-MAR-2002 (first entry)
XX
DE *Drosophila melanogaster* genomic polynucleotide SEQ ID NO 1543.
XX
KW *Drosophila*; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS *Drosophila melanogaster*.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PERKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from *Drosophila* and for elucidating cell signalling and cell-cell
XX interactions.
XX
PS Claim 1; SEQ ID NO 1543; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL20511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABBS72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

50 Sequence 4169 BP; 1248 A; 859 C; 919 G; 1143 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 4; Length 4169;
Best Local Similarity 88.0%; Pred. No. 3.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCAAATTCAGATA 25
DB 3426 AAAAAAAAAACTTACATTTCAGATA 3402

RESULT 7
ABK39977/c
ID ABK39977 standard; DNA; 5933 BP.
XX
AC ABK39977;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #30 strand 1.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytosstatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PE 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 59; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM 000693), CYP11A (NM 000781), CYP11B1 (NM 000497), CYP3A3
CC (NM 000776 and NM 017460), DPYD (NM 000110), EPHX2 (NM 001979), OCLN
CC (NM 002538), TXNRD1 (NM 003330), UGT8 (NM 003360), MRP (NM 004896,
CC NM 019900, NM 019901, NM 019902, NM 019862, NM 019898, NM 019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

50 Sequence 5933 BP; 1686 A; 141 C; 1346 G; 2760 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 6; Length 5933;
Best Local Similarity 88.0%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCAAATTCAGATA 25
DB 2331 AAAAAAAAAATCCCAATTCAGATA 2307

RESULT 8
ABK39978
ID ABK39978 standard; DNA; 5933 BP.
XX
AC ABK39978;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #30 strand 2.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytosstatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PE 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 60; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM 000693), CYP11A (NM 000781), CYP11B1 (NM 000497), CYP3A3
CC (NM 000776 and NM 017460), DPYD (NM 000110), EPHX2 (NM 001979), OCLN
CC (NM 002538), TXNRD1 (NM 003330), UGT8 (NM 003360), MRP (NM 004896,
CC NM 019900, NM 019901, NM 019902, NM 019862, NM 019898, NM 019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 5933 BP; 1498 A; 141 C; 1403 G; 2891 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 6; Length 5933;
 Best Local Similarity 88.0%; Pred. No. 3.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
 DB 3603 AAAAAAAAAAGTTCATTCAGATA 3627

RESULT 9
 ABQ67050/c
 ID ABQ67050 standard; DNA; 6486 BP.
 AC
 XX ABQ67050;
 XX
 DT 28-AUG-2002 (first entry)
 XX
 DE Human angiogenesis associated polynucleotide SEQ ID NO 80.
 XX

KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; antileucoderms;
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;
 KW antineuritic; antirheumatic; antidiabetic; antiporiatic;
 KW antileukosclerotic; ds.
 XX
 OS Homo sapiens.
 XX

PN WO200246454-A2.
 XX
 PD 13-JUN-2002.
 XX
 PE 06-DEC-2001; 2001WO-EP014320.
 XX
 PR 06-DEC-2000; 2000DE-01061338.
 XX

PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Schacht O;
 XX
 DR WPI; 2002-500450/53.
 XX
 PT New nucleic acid fragments for chemically treated angiogenesis-
 associated genes, useful for determining methylation status, e.g. in
 PT diagnosis or treatment of cancer.
 XX
 PS Claim 1; SEQ ID NO 80; 41pp + Sequence Listing; German.
 XX

CC The invention relates to a nucleic acid (I) comprising a segment of 18
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
 CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
 CC related oligomers, are used to evaluate the methylation status and/or
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
 CC Crohn's disease. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6486 BP; 1713 A; 176 C; 1443 G; 3154 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 6486;
 Best Local Similarity 88.0%; Pred. No. 3.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25
 DB 6080 AAAAAAAAAAGTTCATTCAGATA 6056

RESULT 10
 AAS46438/c
 ID AAS46438 standard; DNA; 9510 BP.
 AC
 XX AAS46438;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #160.
 XX

KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
 KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX

PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PE 15-MAR-2001; 2001WO-EP002955.
 XX

PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX

PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumor suppressor
 PT genes and oncogenes, useful in designing primers and probes for analyzing
 PT diseases associated with cytosine methylation state e.g. cancer.
 XX
 PS Claim 1; SEQ ID NO 160; 27pp; English.
 XX

CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and oncogenes
 CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
 CC 500 are missing from the sequence listing) sequences (Ss) and sequences
 CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
 CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
 CC probes for detecting the cytosine methylation state and/or single
 CC nucleotide polymorphisms and also to be used in an array for analysing
 CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
 CC probes can also be used in a method for ascertaining genetic and/or
 CC epigenetic parameters for the diagnosis and/or therapy of existing
 CC diseases or the predisposition to specific diseases, by analysing
 CC cytosine methylations. The parameters may be compared to another set of
 CC genetic and/or epigenetic parameters, the differences serving as basis
 CC for diagnosis and/or prognosis events which are disadvantageous to
 CC patients. The present sequence is one of the 533 genomic sequences
 CC derived from tumour suppressor genes and oncogenes. Sequences with even
 CC numbered Seq ID numbers are the complementary sequence of the
 CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
 CC 533, except for those whose partner sequence is missing). Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 9510 BP; 3043 A; 143 C; 1830 G; 4445 T; 0 U; 49 Other;

Query Match 80.8%; Score 20.2; DB 4; Length 9510;
 Best Local Similarity 88.0%; Pred. No. 3.5e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGATA 25

Db 5970 AAAAAAAAAATCCATTAAATA 5946

RESULT 11
ABL34563/c
ID ABL34563 standard; DNA; 9510 BP.
XX
XX ABL34563;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Human metastasis associated gene SEQ ID NO: 116.
DE
XX
XX Metastasis associated gene; cytostatic; gene therapy; cancer;
KM cytosine methylation; gene; ds.
XX
XX Homo sapiens.
OS
XX MO200177376-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 06-APR-2001; 2001WO-EP003970.
PF
XX
XX 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2002-010922/01.
DR
XX
XX New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
XX Claim 1; SEQ ID NO 116; 23pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention
XX
XX Sequence 9510 BP; 3043 A; 143 C; 1830 G; 4445 T; 0 U; 49 Other;
SQ

Query Match 80.8%; Score 20.2; DB 6; Length 9510;
Best Local Similarity 88.0%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTGAGATA 25
Db 5970 AAAAAAAAAATCCATTAAATA 5946

RESULT 12
ADS99824/c
ID ADS99824 standard; DNA; 9510 BP.
XX
XX ADS99824;
AC
XX
XX 02-DEC-2004 (first entry)
DT
XX
XX Complement of bisulphite treated metastasis-associated human gene #58.
DE
XX
XX Human; ds; gene; Bisulphite; metastasis; cancer; cytostatic;
KM DNA methylation; matrix-assisted laser desorption/ionisation; MALDI;
XX electrospray; mass spectrometry; CpG dinucleotide; solid tumour.
XX
XX Homo sapiens.
OS

EN US2003148327-A1.
XX
XX 07-AUG-2003.
PD
XX
XX 21-JAN-2003; 2003US-00240485.
PE
XX
XX 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
PR 06-APR-2001; 2001WO-EP003970.
XX
XX (OLEK/) OLEK A.
PA (PIEP/) PIEPENBROCK C.
PA (BERL/) BERLIN K.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2002-010922/01.
DR
XX
XX New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
XX Claim 1; SEQ ID NO 116; 9pp; English.
PS
XX
XX The invention relates to a nucleic acid comprising at least 18 bases from
CC a segment of the chemically pretreated DNA of genes associated with
CC metastasis, i.e. any of ADS99709-ADS99906 human genomic sequences or any
CC of the 19 sequences appearing as ADS99911-ADS99929. SEQ ID 2,4,6 etc are
CC the complements of SEQ ID 1,3,5, etc. Also included are an oligomer
CC (particularly an oligonucleotide or peptide nucleic acid) comprising at
CC least one base sequence of at least 9 bases which hybridises to (or is
CC identical with) the sequences referred to above, producing an array of
CC the oligomers on a carrier, obtaining genetic and/or epigenetic
CC parameters for diagnosis and/or therapy of diseases (or predisposition to
CC them) by analysis of cytosine methylation and a kit comprising a
CC bisulphite (disulphite or hydrogen sulphite) and the oligomers. In the
CC method of above 5-unmethylated cytosines in a genomic DNA sample are
CC converted chemically to uracil, or another base with hybridisation
CC properties different from those of cytosine, then fragments of the
CC treated DNA amplified (particularly by polymerase chain reaction) using
CC the oligomers and a polymerase (preferably heat stable) to produce
CC labelled amplicons. These are tested for hybridisation to an array of
CC oligomers and any hybridisation detected. The amplicons are labelled with
CC fluorescent or radioactive markers, or with a detachable mass marker to
CC allow their detection by mass spectrometry, specifically using the matrix
CC -assisted laser desorption/ionisation (MALDI) or electrospray techniques.
CC To improve detection in the mass spectrometer, fragments formed in the
CC instrument have only a single net charge (positive or negative). The
CC genomic DNA is from e.g. a cell line, biopsy sample, blood, or paraffin-
CC embedded tissue sample. Oligonucleotides or peptide-nucleic acids that
CC are complementary to (or identical with) parts of the nucleic acids listed
CC above may be used as primers for amplification of the nucleic acids or
CC their complements, and for determining cytosine methylation status and/or
CC single nucleotide polymorphisms in metastasis-related genes. They can be
CC used for analysis of diseases associated with methylation of CpG
CC dinucleotides and to determine (epi)genetic parameters for diagnosis
CC and/or therapy of disease (or predisposition). The genomic DNA sequences
CC are useful for diagnosis and therapy of solid tumours and cancer. The
CC present sequence is the complementary sequence to a bisulphite treated
CC human gene associated with metastasis. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030148327.
XX
XX Sequence 9510 BP; 3043 A; 143 C; 1830 G; 4445 T; 0 U; 49 Other;
SQ

Query Match 80.8%; Score 20.2; DB 7; Length 9510;
Best Local Similarity 88.0%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTGAGATA 25

Db 5970 AAAAAAAAAATTCCAATTAAATA 5946

RESULT 13
ABL34432
ID ABL34432 standard; DNA; 10151 BP.

XX ABL34432;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2405.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
KM antiarteriosclerotic; antihaemic; cyostatic; noctropic;
KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.

XX Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PE 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

PS Claim 1; SEQ ID NO 2405; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention

XX Sequence 10151 BP; 2631 A; 188 C; 2203 G; 5127 T; 0 U; 2 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 10151;
Best Local Similarity 88.0%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
DB 7821 AAAAAAAAAAGTTTAATTAGATA 7845

RESULT 14
ABL34433/C
ID ABL34433 standard; DNA; 10151 BP.

XX ABL34433;

DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 2406.

DE Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antihaemic; cyostatic; noctropic;
KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.

XX Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PE 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

PS Claim 1; SEQ ID NO 2406; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention

XX Sequence 10151 BP; 3017 A; 188 C; 2298 G; 4646 T; 0 U; 2 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 10151;
Best Local Similarity 88.0%; Pred. No. 3.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCAGATA 25
DB 2331 AAAAAAAAAAATTCCAATTCAATA 2307

RESULT 15
AAS26800
ID AAS26800 standard; DNA; 12542 BP.

XX AAS26800;

DT 07-NOV-2001 (first entry)

DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1774.

XX Human; immunosuppressive; antiarthritic; ds; antirheumatic; cyostatic;
KM cardiac; vasotropic; cerebroprotective; noctropic; neuroprotective;
KM antibacterial; virucide; fungicide; ophthalmological; vulnary;
KM secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KM cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KM cerebral ischaemia; angiogenesis; nervous system disorder;
KM Alzheimer's disease; infection; ocular disorder; cornea infection;
KM wound healing; epithelial cell proliferation; skin ageing; food additive;

XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-488783/53.
 XX
 DR New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 PS Disclosure; SEQ ID NO 1774; 980bp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence is a genomic DNA encoding a partial novel secreted protein of
 CC the invention. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format

Query Match 80.8%; Score 20.2; DB 4; Length 12542;
 Best Local Similarity 88.0%; Pred. No.3.6e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTGAGATA 25
 DB 4239 AAAAAAAAAAGTTCATTGAGATA 4263

RESULT 16
 ABX74149
 ID ABX74149 standard; DNA; 12542 BP.
 XX
 AC ABX74149;
 XX
 DT 19-MAR-2003 (first entry)
 XX
 DE Human novel polynucleotide #977.
 XX
 KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antitateriosclerotic.
 XX
 OS Homo sapiens.
 XX
 PN US2002132753-A1.
 XX
 PD 19-SEP-2002.
 XX
 PF 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.
 XX 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 30-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234233P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236337P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249259P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBEN/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2003-147444/14.
 XX
 KW New polypeptides and nucleic acids, useful in gene therapy for treating,
 KW inhibiting or preventing e.g. neural, immune system, muscular,
 KW respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 KW renal disorders.
 XX
 PS Disclosure; SEQ ID NO 1774; 402bp; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,

CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Epstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABX7173-ABX74167 represent
CC human novel polynucleotides of the invention
XX

SQ Sequence 12542 BP; 3802 A; 2834 C; 2642 G; 3264 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 8; Length 12542;
Best Local Similarity 88.0%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTCAGATA 25
4239 AAAAAAAAAAGTTCATTCAGATA 4263

Db

RESULT 17
AAK78660
ID AAK78660 standard; DNA; 15320 BP.
XX
AC AAK78660;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33472.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184684P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-023637P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0244617P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.

XX Disclosure; SEQ ID NO 33472; 3071bp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytotoxic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK7664 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX

SQ Sequence 15320 BP; 4488 A; 2687 C; 3177 G; 4968 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 4; Length 15320;

Best Local Similarity 88.0%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTCCATTTCAGATA 25
DB 3177 AAAAAAAAAAATACATTTCAGATA 3201

RESULT 18
AAK78662
ID AAK78662 standard; DNA; 15332 BP.
XX
AC AAK78662;
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33474.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; de.
XX
OS Homo sapiens.
PN WO200157182-A2.
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 16-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 01-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.

PR	14-SEP-2000	2000US-0222337
PR	14-SEP-2000	2000US-0223358
PR	14-SEP-2000	2000US-0223399
PR	14-SEP-2000	2000US-022400P
PR	14-SEP-2000	2000US-022401P
PR	14-SEP-2000	2000US-022401P
PR	14-SEP-2000	2000US-0223063P
PR	14-SEP-2000	2000US-0223064P
PR	14-SEP-2000	2000US-0223065P
PR	21-SEP-2000	2000US-0224223P
PR	21-SEP-2000	2000US-0224274P
PR	25-SEP-2000	2000US-0224997P
PR	25-SEP-2000	2000US-0224998P
PR	26-SEP-2000	2000US-0225484P
PR	27-SEP-2000	2000US-0225834P
PR	27-SEP-2000	2000US-0225836P
PR	29-SEP-2000	2000US-0226337P
PR	29-SEP-2000	2000US-0226367P
PR	29-SEP-2000	2000US-0226368P
PR	29-SEP-2000	2000US-0226369P
PR	29-SEP-2000	2000US-0226370P
PR	02-OCT-2000	2000US-0226802P
PR	02-OCT-2000	2000US-0227037P
PR	02-OCT-2000	2000US-0227038P
PR	02-OCT-2000	2000US-0227039P
PR	02-OCT-2000	2000US-0227040P
PR	13-OCT-2000	2000US-0229355P
PR	13-OCT-2000	2000US-0229393P
PR	20-OCT-2000	2000US-0229397P
PR	20-OCT-2000	2000US-0229406P
PR	20-OCT-2000	2000US-02211231P
PR	20-OCT-2000	2000US-02211785P
PR	20-OCT-2000	2000US-02211786P
PR	20-OCT-2000	2000US-02211787P
PR	20-OCT-2000	2000US-0221808P
PR	20-OCT-2000	2000US-0221809P
PR	20-OCT-2000	2000US-0221826P
PR	01-NOV-2000	2000US-0224617P
PR	01-NOV-2000	2000US-0224647P
PR	08-NOV-2000	2000US-02246475P
PR	08-NOV-2000	2000US-02246476P
PR	08-NOV-2000	2000US-02246477P
PR	08-NOV-2000	2000US-02246478P
PR	08-NOV-2000	2000US-02246532P
PR	08-NOV-2000	2000US-02246524P
PR	08-NOV-2000	2000US-02246525P
PR	08-NOV-2000	2000US-02246526P
PR	08-NOV-2000	2000US-02246527P
PR	08-NOV-2000	2000US-02246528P
PR	08-NOV-2000	2000US-02246532P
PR	08-NOV-2000	2000US-02246534P
PR	08-NOV-2000	2000US-02246609P
PR	08-NOV-2000	2000US-02246610P
PR	08-NOV-2000	2000US-02246611P
PR	08-NOV-2000	2000US-02246613P
PR	17-NOV-2000	2000US-0229207P
PR	17-NOV-2000	2000US-0229208P
PR	17-NOV-2000	2000US-0229209P
PR	17-NOV-2000	2000US-0229210P
PR	17-NOV-2000	2000US-0229211P
PR	17-NOV-2000	2000US-0229212P
PR	17-NOV-2000	2000US-0229213P
PR	17-NOV-2000	2000US-0229214P
PR	17-NOV-2000	2000US-0229215P
PR	17-NOV-2000	2000US-0229216P
PR	17-NOV-2000	2000US-0229217P
PR	17-NOV-2000	2000US-0229218P
PR	17-NOV-2000	2000US-0229244P
PR	17-NOV-2000	2000US-0229245P
PR	17-NOV-2000	2000US-0229246P
PR	17-NOV-2000	2000US-0229265P
PR	17-NOV-2000	2000US-0229287P
PR	17-NOV-2000	2000US-0229289P
PR	17-NOV-2000	2000US-0229300P
PR	01-DEC-2000	2000US-0225016P
PR	01-DEC-2000	2000US-0225031P

PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251865P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI: 2001-483426/52.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	useful for preventing, diagnosing and/or treating cancers and metastasis.	
XX		
PS	Disclosure: SEQ ID NO 33474; 3071pp + Sequence Listing; English.	
XX		
CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I), by inserting the	
CC	nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/haematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK62169	
CC	represent sequences used in the exemplification of the present invention	
SO		
XX	Sequence 15332 BP; 4490 A; 2686 C; 3180 G; 4976 T; 0 U; 0 Other;	
Query Match	80.8%; Score 20.2; DB 4; Length 15332;	
Best Local Similarity	88.0%; Pred. No. 3.6e+02;	
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0.		
QY	1 AAAAAAAAAAGTTCCATTTCAGATA 25	
DB	3189 AAAAAAAAAAATAACATTTCAGATA 3213	
RESULT 19		
ID	ACA19757 standard; DNA; 20142 BP.	
AC	ACA19757;	
XX		
DT	19-JUN-2003 (first entry)	
XX		
DE	Prokaryotic essential gene #1414.	
XX		
KM	Antisense; ds; prokaryotic essential gene; cell proliferation;	
KM	drug design; gene.	
XX		
OS	Staphylococcus aureus.	
XX		
PN	WO200277183-A2.	
XX		
PD	03-OCT-2002.	
XX		
PF	21-MAR-2002; 2002WO-US009107.	
XX		

PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JM;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR P-PSDB; ABU15887.
XX
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 7627; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: the sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 20142 BP; 8529 A; 3619 C; 3611 G; 4383 T; 0 U; 0 Other;
SQ
XX
XX Query Match 80.8%; Score 20.2; DB 8; Length 20142;
Best Local Similarity 88.0%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTTCATTTCAGATA 25
DB 2398 AACAAACAGGTTCCATTTCAGATA 2422

RESULT 20
ACN43994/C
ID ACN43994 standard; DNA; 109586 BP.
XX
XX ACN43994;
AC
XX
XX 18-NOV-2004 (first entry)
DT
XX
XX Human genomic sequence hCG23847.
DE
XX

KM Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX NC02003073826-A2.
FN
XX
XX 12-SEP-2003.
PD
XX
XX 28-FEB-2003; 2003MO-US006235.
PF
XX
XX 01-MAR-2002; 2002US-00087192.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
XX Morris DW;
PI
XX
XX WPI; 2003-328604/31.
DR
XX
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
PT
XX
XX Claim 1; SEQ ID NO 220; 0bp; English.
PS
XX
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
XX Sequence 109586 BP; 31351 A; 20631 C; 21224 G; 35318 T; 0 U; 1062 Other;
SQ
XX
XX Query Match 80.8%; Score 20.2; DB 11; Length 109586;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTTCATTTCAGATA 25
DB 97181 AAAAAAAAAAGTTCATTTCAGATA 97157

RESULT 21
AAA70156
ID AAA70156 standard; DNA; 7491 BP.
XX
XX AAA70156;
AC
XX
XX 07-NOV-2000 (first entry)
DT
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:289.
DE
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalaria; malaria; protozoacide; infection; insecticide; de.
XX
XX Plasmodium falciparum.
OS
XX
XX MO200025728-A2.
PN
XX
XX 11-MAY-2000.
PD
XX
XX 05-NOV-1999; 99MO-US026796.
PF
XX
XX 05-NOV-1998; 98US-0107131P.
PR
XX
XX (HOFF/) HOFFMAN S.
PA

PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
DR
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection.
XX
PS Disclosure; Page 496-497; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
CC (II) are useful for the development of vaccines against P. falciparum
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with P. falciparum. Furthermore, (I) (especially when they
CC are rifins or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent P. falciparum infection, or they can be used
CC to identify drug resistance in P. falciparum. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX
SQ Sequence 7491 BP; 3252 A; 787 C; 964 G; 2488 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 3; Length 7491;
Best Local Similarity 91.3%; Pred. No. 5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAAAGTTCATTCAGA 23
DB 3895 AATAAAAAAGTCTTAATTCAGA 3917
XX
RESULT 22
AAL25065/C
ID AAL25065 standard; cDNA; 665 BP.
XX
AC AAL25065;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 17522.
XX
KM Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN W0200151628-A2.
XX
PD 19-JUL-2001.
XX
PE 10-JAN-2001; 2001MO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.

PR 25-JUL-2000; 2000US-0220534P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI; 2001-451856/48.
DR
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
PT
XX
PS Claim 1; Page 3238; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterizing treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 665 BP; 161 A; 152 C; 134 G; 218 T; 0 U; 0 Other;
XX
Query Match 77.6%; Score 19.4; DB 4; Length 665;
Best Local Similarity 95.2%; Pred. No. 6.6e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAAAGTTCATTCAGA 21
DB 502 AAAAAAAAAAGTTCATTCAGA 482
XX
RESULT 23
AAC55139/C
ID AAC55139 standard; DNA; 42 BP.
XX
AC AAC55139;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 80176.
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway; metabolic pathway;
KM promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX
PN EP1033405-A2.
XX
PN
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129645P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.

PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147933P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148655P.
PR 13-AUG-1999; 99US-0148684P.
PR 15-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 18-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.

```
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 76.8%; Score 19.2; DB 3; Length 42;
Best Local Similarity 87.5%; Pred. No. 7.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCCATTCAGAT 24
   |||||
Db 35 AAAAAAAAAATTCAATTGAGAT 12

RESULT 24
AAK72039/c
ID AAK72039 standard; DNA; 424 BP.
XX
AC AAK72039;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26851.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
```

```
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
```

PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 03-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Disclosure; SEQ ID NO 26851; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 424 BP; 163 A; 57 C; 69 G; 135 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 4; Length 424;
 Best Local Similarity 87.5%; Pred. No. 7.7e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCAAATTCAGAT 24
 DB 414 AAAAAAAAAAGTCCCTCTGAT 391
 RESULT 25
 ID ABV10721 standard; cDNA; 533 BP.
 XX
 AC ABV10721;
 XX
 DT 13-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 10712.
 XX
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207445P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JB;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 1731; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 533 BP; 150 A; 101 C; 111 G; 169 T; 0 U; 2 Other;
 Query Match 76.8%; Score 19.2; DB 5; Length 533;
 Best Local Similarity 87.5%; Pred. No. 7.8e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCAAATTCAGAT 24
 DB 151 AAAAAAAAAAGTCCCAATCAAT 174
 RESULT 26
 ADB61077/C
 ID ADB61077 standard; DNA; 724 BP.

AC ADE61077;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene B1823499, SEQ ID NO 6991.
XX
KW Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
FN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEMO) GEN HOSPITAL CORP.
XX (FARB) BAYER AG.
XX
PI Woolf C, D'Urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; B1823499.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 724 BP; 218 A; 158 C; 146 G; 202 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 10; Length 724;
Best Local Similarity 87.5%; Pred. No. 7.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTTCAGAT 24
Db 648 AAAAAAAAAAGTCCATTTCAGAT 625

RESULT 27
ID ADE60529/c
XX ADE60529 standard; DNA; 724 BP.
XX
AC ADE60529;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene B1823499, SEQ ID NO 6438.
XX
KW Human; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
FN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEMO) GEN HOSPITAL CORP.
XX (FARB) BAYER AG.
XX
PI Woolf C, D'Urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; B1823499.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 724 BP; 218 A; 158 C; 146 G; 202 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 10; Length 724;

Best Local Similarity 87.5%; Pred. No. 7.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTTCATTCAGAT 24
DB 648 AAAAAAAAAAGTTCATTCAGAT 625

RESULT 28
ABV11370
ID ABV11370 standard; cDNA; 832 BP.
XX
AC ABV11370;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker CDNA 11361.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
PT WPI; 2001-662795/76.
XX
DR
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 1849; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 832 BP; 273 A; 173 C; 163 G; 219 T; 0 U; 4 Other;
Query Match 76.8%; Score 19.2; DB 5; Length 832;
Best Local Similarity 87.5%; Pred. No. 7.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTTCATTCAGAT 24
DB 595 AAAAAAAAAAGTTCATTCAT 618
RESULT 29

AAC59496/c
ID AAC59496 standard; cDNA; 1356 BP.
XX
AC AAC59496;
XX
DT 26-JAN-2001 (first entry)
XX
DE Human secreted protein gene 48 SEQ ID NO:58.
XX
KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW infection; ocular disorder; wound healing; skin aging; food additive;
KW preservative; ss.
XX
OS Homo sapiens.
XX
PN WO200056755-A1.
XX
PD 28-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US006830.
XX
PR 19-MAR-1999; 99US-0125361P.
PR 10-DEC-1999; 99US-0169910P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsculis G;
XX
PT WPI; 2000-587661/55.
XX
DR P-PSDB; AAB34139.
XX
PT New isolated nucleic acid molecules encoding 49 human secreted proteins
PT used for preventing, treating or ameliorating medical conditions, for
PT diagnosing pathological conditions or as food additives or preservatives.
XX
PS Claim 1; Page 362-363; 419pp; English.
XX
CC The polynucleotide sequences given in AAC59449 to AAC59497 encode the
CC human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
CC AAB34216 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissue
CC and cells the genes are expressed in. Examples of activities include:
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC and vulnerary. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
CC cancer of the breast or liver, cardiovascular disorders, cerebrovascular
CC disorders, angiogenesis, nervous system disorders, infections caused by
CC bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 1356 BP; 444 A; 248 C; 303 G; 361 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 3; Length 1356;
Best Local Similarity 87.5%; Pred. No. 8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTTCATTCAGATA 25
DB 203 AACAAGAGTTCATTCAGATA 180

RESULT 30

ACL34950/C
ID ACL34950 standard; cDNA, 1654 BP.

ACL34950;

02-JUN-2005 (first entry)

DE Rice stress-regulated promoter SEQ ID NO:13513.

KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;
KM agriculture.

OS Oryza sativa.

PN MO2003008540-A2.

PD 30-JAN-2003.

PF 21-JUN-2002; 2002WO-US019668.

PR 22-JUN-2001; 2001US-0300112P.

PR 24-AUG-2001; 2001US-0314662P.

PR 26-SEP-2001; 2001US-0325277P.

PR 21-NOV-2001; 2001US-0332132P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Kieps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;

PI Moughamer T, Provart N, Ricke D, Zhu T;

XX WPI; 2003-248011/24.

PS New stress-responsive nucleic acid, useful for altering the

PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold

XX stress, salt stress or osmotic stress.

PS Claim 48; SEQ ID NO 13513; 89pp; English.

XX The invention relates to novel abiotic stress responsive polynucleotides

CC and polypeptides. Also disclosed are vectors, expression cassettes, host

CC cells, and plants containing such polynucleotides. Also disclosed are

CC methods for using the polynucleotides and polypeptides to alter the

CC responsiveness of a plant to abiotic stress. The invention is useful in

CC agriculture. The nucleic acid is useful for determining whether a test

CC plant has been exposed to an abiotic stress condition. It is also useful

CC for selecting an agent that alters abiotic stress regulated

CC polynucleotide expression in a plant cell, and to identify a homolog or

CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid

CC molecule and the polypeptide encoded by it are useful in altering the

CC responsiveness of a plant to an abiotic stress, such as cold stress, salt

CC stress, osmotic stress or any of their combinations. The present sequence

CC is used in the exemplification of the invention

XX

SO Sequence 1654 BP; 535 A; 312 C; 296 G; 503 T; 0 U; 8 Other;

Query Match 76.8%; Score 19.2; DB 11; Length 1654;

Best Local Similarity 87.5%; Pred. No. 8e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24

DB 406 AAAAAAAAAAGTTCAACTCAGTT 383

ID ACL36184 standard; cDNA; 2000 BP.

ACL36184;

02-JUN-2005 (first entry)

DE Rice stress-regulated promoter SEQ ID NO:14747.

KW ss; abiotic stress tolerance; transgenic plant; plant; cereal;

KM agriculture.

OS Oryza sativa.

PN MO2003008540-A2.

PD 30-JAN-2003.

PF 21-JUN-2002; 2002WO-US019668.

PR 22-JUN-2001; 2001US-0300112P.

PR 24-AUG-2001; 2001US-0314662P.

PR 26-SEP-2001; 2001US-0325277P.

PR 21-NOV-2001; 2001US-0332132P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Kieps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;

PI Moughamer T, Provart N, Ricke D, Zhu T;

XX WPI; 2003-248011/24.

PS New stress-responsive nucleic acid, useful for altering the

PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold

XX stress, salt stress or osmotic stress.

PS Claim 48; SEQ ID NO 14747; 89pp; English.

XX The invention relates to novel abiotic stress responsive polynucleotides

CC and polypeptides. Also disclosed are vectors, expression cassettes, host

CC cells, and plants containing such polynucleotides. Also disclosed are

CC methods for using the polynucleotides and polypeptides to alter the

CC responsiveness of a plant to abiotic stress. The invention is useful in

CC agriculture. The nucleic acid is useful for determining whether a test

CC plant has been exposed to an abiotic stress condition. It is also useful

CC for selecting an agent that alters abiotic stress regulated

CC polynucleotide expression in a plant cell, and to identify a homolog or

CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid

CC molecule and the polypeptide encoded by it are useful in altering the

CC responsiveness of a plant to an abiotic stress, such as cold stress, salt

CC stress, osmotic stress or any of their combinations. The present sequence

CC is used in the exemplification of the invention

XX

SO Sequence 2000 BP; 610 A; 410 C; 374 G; 602 T; 0 U; 4 Other;

Query Match 76.8%; Score 19.2; DB 11; Length 2000;

Best Local Similarity 87.5%; Pred. No. 8.1e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24

DB 1202 AATTAATTAATTCATTCAGAT 1225

RESULT 32

ABL28464

ABL28464 standard; DNA; 3012 BP.

AC ABL28464;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 36865.

KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li FWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 36865; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB161675) and the encoded proteins (AB57737-
 CC AB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 3012 BP; 913 A; 589 C; 663 G; 847 T; 0 U; 0 Other;
 XX
 Query Match 76.8%; Score 19.2; DB 4; Length 3012;
 Best Local Similarity 87.5%; Pred. No. 8.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCATTCCAGAT 24
 Db 600 AAAAAAAAAAGTCCATTCCAGAT 623
 XX
 RESULT 33
 ABA05542
 ID ABA05542 standard; DNA; 3453 BP.
 XX
 AC ABA05542;
 XX
 DT 25-FEB-2002 (first entry)
 XX
 DE Murine alpha1B adrenergic receptor promoter sequence.
 XX
 KM Mouse; alpha1B adrenergic receptor; antiparkinsonian; transgenic animal;
 KM alpha1A; alpha1D; neurodegenerative disorder; cardiovascular disorder;
 KM Parkinson's disease; ds.
 XX
 OS Mus sp.
 XX
 PN WO200184919-A1.
 XX
 PD 15-NOV-2001.
 XX
 PF 04-MAY-2001; 2001WO-US014577.
 XX
 PR 10-MAY-2000; 2000US-00568255.
 XX
 PA (CLEV-) CLEVELAND CLINIC FOUND.

XX
 PI Perez DM, Zusck M;
 XX
 DR WPI; 2002-049379/06.
 XX
 PT Novel transgenic non-human mammalian animal useful for screening drugs,
 PT comprises transgene encoding exogenous wild-type or constitutively-active
 PT mutant alpha1A, alpha1B or alpha1D adrenergic receptor.
 XX
 PS Example 1; Fig 2; 47pp; English.
 XX
 CC The invention relates to a transgenic animal having integrated within its
 CC genome a transgene encoding an exogenous wild-type or constitutively-
 CC active mutant alpha1A, alpha1B or alpha1D adrenergic receptor. The
 CC transgene is operably linked to a promoter that drives expression of the
 CC transgene in cells innervated by the sympathetic nervous system, and the
 CC transgenic animal exhibits an abnormal phenotype. The transgenic animal
 CC is useful for screening for a compound which modulates function of
 CC alpha1BAR, by administering the compound to the animal, and assaying for
 CC changes in the abnormal phenotype. The animal exhibits neurodegenerative
 CC symptoms or symptoms of a cardiovascular disorder, and the assay involves
 CC assaying for an improvement in or delay in progression of the symptoms,
 CC or evaluating the locomotor activity of the animal. The transgenic animal
 CC is also useful for screening a drug for activity against a
 CC neurodegenerative disorder, such as Parkinson's disease, or a
 CC cardiovascular disorder. The present sequence is the promoter of the
 CC murine alpha1B adrenergic receptor
 XX
 SQ Sequence 3453 BP; 957 A; 898 C; 788 G; 810 T; 0 U; 0 Other;
 XX
 Query Match 76.8%; Score 19.2; DB 6; Length 3453;
 Best Local Similarity 87.5%; Pred. No. 8.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGTCCATTCCAGAT 24
 Db 2335 AAAAAAAAAAGTCCATTCCATAT 2358
 XX
 RESULT 34
 ABL32684/C
 ID ABL32684 standard; DNA; 6392 BP.
 XX
 AC ABL32684;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 657.
 XX
 KM Human; immune system disease; cytosine methylation; antiasthmatic;
 KM antiarteriosclerotic; antihaemic; cytotoxic; neurotropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; antiarthritic; antidiabetic; antiparasitic;
 KM antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KM ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
XX
PS Claim 1; SEQ ID NO 657; 32pp + Sequence Listing; German.
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6392 BP; 1743 A; 170 C; 1443 G; 3036 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 6; Length 6392;
Best Local Similarity 87.5%; Pred. No. 8.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCATTTCAGAT 24
DB 1234 AAAAAAAAACTCCATTTCAGAT 1211
RESULT 35
ABL34506/C
ID ABL34506 standard; DNA; 6392 BP.
XX
AC ABL34506;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human metastasis associated gene SEQ ID NO: 59.
XX
KM Metastasis associated gene; cytostatic; gene therapy; cancer;
KM cytosine methylation; gene; de.
OS
OS Homo sapiens.
XX
PN WO200177376-A2.
XX
PD 18-OCT-2001.
XX
PE 06-APR-2001; 2001WO-EP003970.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIDENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-010922/01.
DR
XX New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
PS Claim 1; SEQ ID NO 59; 23pp + Sequence Listing; English.
CC The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention
XX
SQ Sequence 6392 BP; 1743 A; 170 C; 1443 G; 3036 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 6; Length 6392;
Best Local Similarity 87.5%; Pred. No. 8.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCATTTCAGAT 24
DB 1234 AAAAAAAAACTCCATTTCAGAT 1211
RESULT 36
ADS99767/C
ID ADS99767 standard; DNA; 6392 BP.
XX
AC ADS99767;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bisulphite treated human gene associated with metastasis #30.
XX
KM Human; de; gene; Bisulphite; metastasis; cancer; cytostatic;
KM DNA methylation; matrix-assisted laser desorption/ionisation; MALDI;
KM electrospray; mass spectrometry; Cpg dinucleotide; solid tumour.
XX
OS Homo sapiens.
XX
PN US2003148327-A1.
XX
PD 07-AUG-2003.
XX
PE 21-JAN-2003; 2003US-00240485.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
PR 06-APR-2001; 2001WO-EP003970.
XX
PA (OLEK/) OLEK A.
PA (PIEP/) PIEPENBROCK C.
PA (BERL/) BERLIN K.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010922/01.
XX
PT New nucleic acid derived from chemically treated metastasis genes, useful
PT for diagnosis of cancers by analysis of cytosine methylation, also for
PT treatment.
XX
PS Claim 1; SEQ ID NO 59; 9pp; English.
XX
XX The invention relates to a nucleic acid comprising at least 18 bases from
CC a segment of the chemically pretreated DNA of genes associated with
CC metastasis, i.e. any of ADS99709-ADS99906 human genomic sequences or any
CC of the 19 sequences appearing as ADS99911-ADS99929. SEQ ID 2,4,6 etc are
CC the complements of SEQ ID 1,3,5, etc. Also included are an oligomer
CC (particularly an oligonucleotide or peptide nucleic acid) comprising at
CC least one base sequence of at least 9 bases which hybridises to (or is
CC identical with) the sequences referred to above, producing an array of
CC the oligomers on a carrier, obtaining genetic and/or epigenetic
CC parameters for diagnosis and/or therapy of diseases (or predisposition to
CC them) by analysis of cytosine methylation and a kit comprising a
CC bisulphite (disulphite or hydrogen sulphite) and the oligomers. In the
CC method of above 5-unmethylated cytosines in a genomic DNA sample are
CC converted chemically to uracil, or another base with hybridisation
CC properties different from those of cytosine, then fragments of the
CC treated DNA amplified (particularly by polymerase chain reaction) using
CC the oligomers and a polymerase (preferably heat stable) to produce
CC labelled amplicons. These are tested for hybridisation to an array of
CC oligomers and any hybridisation detected. The amplicons are labelled with
CC fluorescent or radioactive markers, or with a detachable mass marker to
CC allow their detection by mass spectrometry, specifically using the matrix
CC assisted laser desorption/ionisation (MALDI) or electrospray techniques.

CC To improve detection in the mass spectrometer, fragments formed in the
CC instrument have only a single net charge (positive or negative). The
CC genomic DNA is from e.g. a cell line, biopsy sample, blood, or paraffin-
CC embedded tissue sample. Oligonucleotides or peptide-nucleic acids listed
CC are complementary to (or identical with) parts of the nucleic acids listed
CC above may be used as primers for amplification of the nucleic acids or
CC their complements, and for determining cytosine methylation status and/or
CC single nucleotide polymorphisms in metastasis-related genes. They can be
CC used for analysis of diseases associated with methylation of CpG
CC dinucleotides and to determine (epi)genetic parameters for diagnosis
CC and/or therapy of disease (or predisposition). The genomic DNA sequences
CC are useful for diagnosis and therapy of solid tumours and cancer. The
CC present sequence is a bisulphite treated human gene associated with
CC metastasis. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?docID=20030148327.
XX

SO Sequence 6392 BP; 1743 A; 170 C; 1443 G; 3036 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 7; Length 6392;
Best Local Similarity 87.5%; Pred. No. 8.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTTCAGAT 24
DB 1234 AAAAAAAAAACCTTCATTTCMAAT 1211

RESULT 37
ABL33421/c
ID ABL33421 standard; DNA; 9707 BP.
XX
AC ABL33421;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1394.
XX
XX Human immune system associated gene; cytosine methylation; antiasthmatic;
KW antiasthmatic; anti-HIV; anticonvulsant; ophthalmological;
KW antineoplastic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
de.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
PN
XX 03-JAN-2002.
PD
XX 02-JUL-2001; 2001WO-EP007537.
PF
XX 30-JUN-2000; 2000DE-01032529.
PR
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
XX Claim 1, SEQ ID NO 1394, 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders.

CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX

SO Sequence 9707 BP; 2786 A; 64 C; 1764 G; 5093 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 6; Length 9707;
Best Local Similarity 87.5%; Pred. No. 8.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTTCAGAT 24
DB 540 AAAAAAAAAATTCCATTTCMAAT 517

RESULT 38
AAK65420/c
ID AAK65420 standard; DNA; 10695 BP.
XX
AC AAK65420;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20232.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosinetic; gene therapy; vaccine; metastasis; de.
OS
XX Homo sapiens.
XX
XX WO200157182-A2.
PN
XX 09-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US001354.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR
XX 04-FEB-2000; 2000US-0180628P.
PR
XX 24-FEB-2000; 2000US-0184664P.
PR
XX 02-MAR-2000; 2000US-0186350P.
PR
XX 16-MAR-2000; 2000US-0189874P.
PR
XX 17-MAR-2000; 2000US-0190076P.
PR
XX 18-APR-2000; 2000US-0198123P.
PR
XX 19-MAY-2000; 2000US-0205515P.
PR
XX 07-JUN-2000; 2000US-0209467P.
PR
XX 28-JUN-2000; 2000US-0214866P.
PR
XX 30-JUN-2000; 2000US-0215135P.
PR
XX 07-JUL-2000; 2000US-0216647P.
PR
XX 07-JUL-2000; 2000US-0216880P.
PR
XX 11-JUL-2000; 2000US-0217487P.
PR
XX 14-JUL-2000; 2000US-0218290P.
PR
XX 26-JUL-2000; 2000US-0220963P.
PR
XX 26-JUL-2000; 2000US-0220964P.
PR
XX 14-AUG-2000; 2000US-0224518P.
PR
XX 14-AUG-2000; 2000US-0224519P.
PR
XX 14-AUG-2000; 2000US-0225213P.
PR
XX 14-AUG-2000; 2000US-0225214P.
PR
XX 14-AUG-2000; 2000US-0225266P.
PR
XX 14-AUG-2000; 2000US-0225267P.
PR
XX 14-AUG-2000; 2000US-0225268P.
PR
XX 14-AUG-2000; 2000US-0225270P.
PR
XX 14-AUG-2000; 2000US-0225447P.
PR
XX 14-AUG-2000; 2000US-0225757P.
PR
XX 14-AUG-2000; 2000US-0225758P.
PR
XX 14-AUG-2000; 2000US-0225759P.
PR
XX 18-AUG-2000; 2000US-0226279P.
PR
XX 22-AUG-2000; 2000US-0226681P.
PR
XX 22-AUG-2000; 2000US-0226686P.
PR
XX 22-AUG-2000; 2000US-0227182P.
PR
XX 23-AUG-2000; 2000US-0227009P.

```
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249259P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 05-DEC-2000; 2000US-0251986P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251859P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM,
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 20232; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 10695 BP; 2515 A; 2871 C; 2419 G; 2890 T; 0 U; 0 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 4; Length 10695;
XX Best Local Similarity 87.5%; Pred. No. 8.5e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX
XX 1 AAAAAAAAAAGTCCATTCCAGAT 24
XX Db 1340 AAAAAAAAAATTCAAGTCAGTT 1317
XX
XX RESULT 39
XX AAK6438/c
XX ID AAK6438 standard; DNA; 18664 BP.
XX
XX
```

AC	AAR64438;
XX	07-NOV-2001 (first entry)
XX	Human immune/haematopoietic
XX	Human; immune; haematopoietic
KW	cyclostatic; gene therapy; vac
OS	Homo sapiens.
XX	MOZ00157182-A2.
PN	09-AUG-2001.
PD	17-JUN-2001; 2001WO-US001354.
PF	31-JAN-2000; 2000US-01790653P.
XX	04-FEB-2000; 2000US-01806282P.
PR	24-FEB-2000; 2000US-01846644P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189674P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205651P.
PR	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214866P.
PR	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.
PR	07-JUL-2000; 2000US-0216880P.
PR	11-JUL-2000; 2000US-0217487P.
PR	11-JUL-2000; 2000US-0217496P.
PR	14-JUL-2000; 2000US-0218590P.
PR	26-JUL-2000; 2000US-0220963P.
PR	26-JUL-2000; 2000US-0220964P.
PR	14-AUG-2000; 2000US-0224518P.
PR	14-AUG-2000; 2000US-0224519P.
PR	14-AUG-2000; 2000US-0225213P.
PR	14-AUG-2000; 2000US-0225214P.
PR	14-AUG-2000; 2000US-0225266P.
PR	14-AUG-2000; 2000US-0225267P.
PR	14-AUG-2000; 2000US-0225268P.
PR	14-AUG-2000; 2000US-0225270P.
PR	14-AUG-2000; 2000US-0225477P.
PR	14-AUG-2000; 2000US-0225574P.
PR	14-AUG-2000; 2000US-0225758P.
PR	14-AUG-2000; 2000US-0225759P.
PR	18-AUG-2000; 2000US-0226279P.
PR	22-AUG-2000; 2000US-0226681P.
PR	22-AUG-2000; 2000US-0226688P.
PR	22-AUG-2000; 2000US-0227182P.
PR	23-AUG-2000; 2000US-0227089P.
PR	30-AUG-2000; 2000US-0228924P.
PR	01-SEP-2000; 2000US-0229287P.
PR	01-SEP-2000; 2000US-0229343P.
PR	01-SEP-2000; 2000US-0229344P.
PR	01-SEP-2000; 2000US-0229345P.
PR	05-SEP-2000; 2000US-0229509P.
PR	05-SEP-2000; 2000US-0229513P.
PR	06-SEP-2000; 2000US-0230437P.
PR	06-SEP-2000; 2000US-0230438P.
PR	08-SEP-2000; 2000US-0231342P.
PR	08-SEP-2000; 2000US-0231343P.
PR	08-SEP-2000; 2000US-0231344P.
PR	08-SEP-2000; 2000US-0231413P.
PR	08-SEP-2000; 2000US-0231414P.
PR	08-SEP-2000; 2000US-0232080P.
PR	12-SEP-2000; 2000US-0231368P.
PR	14-SEP-2000; 2000US-0232397P.
PR	14-SEP-2000; 2000US-0232398P.
PR	14-SEP-2000; 2000US-0232399P.
PR	14-SEP-2000; 2000US-0232400P.

PR	14-SEP-2000	2000US-0232401P
PR	14-SEP-2000	2000US-0233063P
PR	14-SEP-2000	2000US-0233064P
PR	14-SEP-2000	2000US-0233065P
PR	21-SEP-2000	2000US-0234223P
PR	21-SEP-2000	2000US-0234274P
PR	25-SEP-2000	2000US-0234977P
PR	25-SEP-2000	2000US-0234998P
PR	26-SEP-2000	2000US-0235484P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235836P
PR	29-SEP-2000	2000US-0236337P
PR	29-SEP-2000	2000US-0236367P
PR	29-SEP-2000	2000US-0236368P
PR	29-SEP-2000	2000US-0236369P
PR	29-SEP-2000	2000US-0236370P
PR	02-OCT-2000	2000US-0238082P
PR	02-OCT-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	13-OCT-2000	2000US-0237945P
PR	13-OCT-2000	2000US-0239393P
PR	13-OCT-2000	2000US-0239397P
PR	20-OCT-2000	2000US-0240960P
PR	20-OCT-2000	2000US-0241221P
PR	20-OCT-2000	2000US-0241785P
PR	20-OCT-2000	2000US-0241786P
PR	20-OCT-2000	2000US-0241808P
PR	20-OCT-2000	2000US-0241809P
PR	20-OCT-2000	2000US-0241826P
PR	01-NOV-2000	2000US-0244617P
PR	08-NOV-2000	2000US-0244647P
PR	08-NOV-2000	2000US-0246475P
PR	08-NOV-2000	2000US-0246476P
PR	08-NOV-2000	2000US-0246477P
PR	08-NOV-2000	2000US-0246478P
PR	08-NOV-2000	2000US-0246523P
PR	08-NOV-2000	2000US-0246524P
PR	08-NOV-2000	2000US-0246552P
PR	08-NOV-2000	2000US-0246561P
PR	08-NOV-2000	2000US-0246562P
PR	08-NOV-2000	2000US-0246563P
PR	08-NOV-2000	2000US-0246569P
PR	08-NOV-2000	2000US-0246610P
PR	08-NOV-2000	2000US-0246611P
PR	08-NOV-2000	2000US-0246613P
PR	17-NOV-2000	2000US-0249207P
PR	17-NOV-2000	2000US-0249208P
PR	17-NOV-2000	2000US-0249209P
PR	17-NOV-2000	2000US-0249210P
PR	17-NOV-2000	2000US-0249211P
PR	17-NOV-2000	2000US-0249212P
PR	17-NOV-2000	2000US-0249213P
PR	17-NOV-2000	2000US-0249214P
PR	17-NOV-2000	2000US-0249215P
PR	17-NOV-2000	2000US-0249216P
PR	17-NOV-2000	2000US-0249217P
PR	17-NOV-2000	2000US-0249218P
PR	17-NOV-2000	2000US-0249244P
PR	17-NOV-2000	2000US-0249245P
PR	17-NOV-2000	2000US-0249264P
PR	17-NOV-2000	2000US-0249265P
PR	17-NOV-2000	2000US-0249267P
PR	17-NOV-2000	2000US-0249289P
PR	01-DEC-2000	2000US-0249300P
PR	01-DEC-2000	2000US-0250160P
PR	01-DEC-2000	2000US-0250319P
PR	05-DEC-2000	2000US-0251030P
PR	05-DEC-2000	2000US-0251988P
PR	05-DEC-2000	2000US-0256179P
PR	05-DEC-2000	2000US-0256179P

PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246509P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251719P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251987P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254077P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI, 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 20233; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic

CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 18664 BP; 4872 A; 4672 C; 4425 G; 4695 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 4; Length 18664;
Best Local Similarity 87.5%; Pred. No. 8.e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAGTTCGATTCAGAT 24
Db 9309 AAAAAAAAAATTCAGTCAGTT 9286
RESULT 41
ID ABR42519 standard; DNA; 32221 BP.
XX ABR42519;
AC
XX 21-MAY-2002 (first entry)
DT
XX
DE Genomic sequence #418 encoding novel human connective tissue polypeptide.
XX
KW Human; connective tissue related disorder; cancer; gene therapy;
XX cytoskeletal; gene; ds.
KM
XX Homo sapiens.
OS
XX W0200155343-A1.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001322.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-020515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.

RESULT 42
ADB60675
ID ADB60675 standard; DNA; 32221 BP.
XX
AC ADB60675;
XX
DT 04-DEC-2003 (first entry)
XX
DE Connective tissue related genomic DNA #418.
XX
KM cytosolic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;
KM antiarteriosclerotic; immunosuppressive; antithrombotic; antiarthritic;
KM antileukemic; antiallergic; antiasmatic; dermatological;
KM nephrotic; virucide; fungicide; antibacterial; antiparasitic;
KM gene therapy; da; connective tissues disorder; rheumatoid arthritis;
KM systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;
KM cancer metastasis; neoplasia; leukaemia; neurodegenerative disease;
KM Alzheimer's disease; Parkinson's disease; cardiovascular disease;
KM atherosclerosis; myocarditis; cardiopulmonary bypass complication;
KM autoimmune disease; multiple sclerosis; allergic reaction; asthma;
KM rhinitis; eczema; inflammatory condition; Crohn's disease; nephritis;
KM gastrointestinal disorder; inflammatory bowel disease;
KM organ transplant rejection; immune system disorder; Bruton's disease;
KM X-linked lymphoproliferative syndrome;
KM B-cell lymphoproliferative disorder; HIV; AIDS; infection;
KM chromosome identification; chromosome mapping;
KM connective tissue related polynucleotide; gene; da.
XX
OS Homo sapiens.
XX
PN US2003054375-A1.
XX
PD 20-MAR-2003.
XX
PF 07-MAR-2002; 2002US-00092154.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0188874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 23-AUG-2000; 2000US-0227183P.
PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236378P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246533P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.

XX 22-JUL-2004.
XX 22-DEC-2003; 2003WO-US041389.
XX 27-DEC-2002; 2002US-00330773.
XX (SAGR-) SAGRES DISCOVERY INC.
XX Morris DW, Malandro MS;
XX WPI; 2004-543781/52.
XX
XX New isolated cancer associated nucleic acids comprising at least 10
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX cancers such as leukemia and lymphoma.
XX
XX Claim 1; SEQ ID NO 390; 199pp; English.
XX
XX The present invention relates to cancer associated sequences (ADQ97025-
XX ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 90120 BP; 22629 A; 20518 C; 21693 G; 25260 T; 0 U; 20 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 12; Length 90120;
XX Best Local Similarity 87.5%; Pred. No. 9e+02; 3; Indels 0; Gaps 0;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAGTTCGAATTCAGAT 24
XX |||||
XX 84681 AAAAAAAAAATTCATTTCAGAT 84704
XX
XX
XX RESULT 45
XX ABD33448/C
XX ID ABD33448 standard; DNA; 115829 BP.
XX
XX ABD33448;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human cancer-associated (CA) gene HD07-085.
XX
XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX de; cancer; cytostatic.
XX
XX Homo sapiens.
XX
XX WO2004058146-A2.
XX
XX 15-JUL-2004.
XX
XX 15-DEC-2003; 2003WO-US040081.
XX
XX 17-DEC-2002; 2002US-00322281.
XX
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX
XX WPI; 2004-499109/47.
XX
XX Novel human cancer associated protein encoded within open reading frame
XX of cancer associated gene, useful as targets for diagnosing cancer.
XX
XX Claim 16; SEQ ID NO 590; 182pp; English.
XX
XX The invention relates to cancer-associated proteins (CAP) and the cancer-
XX associated (CA) nucleic acids encoding them. The invention also relates

CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 115829 BP; 32585 A; 23364 C; 24898 G; 34704 T; 0 U; 278 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 13; Length 115829;
XX Best Local Similarity 87.5%; Pred. No. 9e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAAGTTCGAATTCAGAT 24
XX |||||
XX 18934 AAAAAAAAAAGTCAATTAAGAT 18911
XX
XX
XX RESULT 46
XX AAC90666/C
XX ID AAC90666 standard; DNA; 3696 BP.
XX
XX AAC90666;
XX
XX 20-MAR-2001 (first entry)
XX
XX Strawberry flowering regulation protein coding sequence SEQ ID NO: 37.
XX
XX Strawberry; flowering regulation; floral homeotic gene; field crop;
XX harvesting; fruit production; de.
XX
XX Fragaria vesca.
XX
XX WO200071722-A1.
XX
XX 30-NOV-2000.
XX
XX 24-MAY-2000; 2000WO-US014297.
XX
XX 25-MAY-1999; 99US-00318789.
XX
XX 24-MAY-2000; 2000US-00318789.
XX
XX (DNAP) DNA PLANT TECHNOLOGY CORP.
XX
XX Oeller P, Guttersen N;
XX
XX WPI; 2001-025165/03.
XX
XX P-PSDB; AAB50269.
XX
XX Novel nucleic acid involved in controlling plant flowering processes is
XX useful for generating transgenic plants, in particular strawberry plants
XX having altered flowering behavior such as early, delayed or day-neutral
XX flowering.
XX
XX Disclosure; Page 82-87; 97pp; English.
XX
XX The present invention provides the nucleic acid and protein sequences of
XX a number of proteins from the strawberry which are involved in the
XX regulation of flowering. These were identified using primers based on the
XX homologous sequences from A. thaliana, B. napus and R. sativus. They can
XX be used in the production of transgenic field crops whose flowering is
XX regulated and the time of fruiting and harvesting can be manipulated
XX
XX Sequence 3696 BP; 1175 A; 618 C; 648 G; 1255 T; 0 U; 0 Other;

Query Match 75.2%; Score 18.8; DB 4; Length 3696;
Best Local Similarity 90.9%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

4 AAAAAAAAAAGTCCAAATTCAGATA 25
AAAAAAAAAGTCCAAATTCAGATA 2905

Db 2926 AAAAAAAAAAGTCCAAATTCAGATA 2905

RESULT 47
ADZ12926
ID ADZ12926 standard; DNA; 235962 BP.
AC ADZ12926;
XX 16-JUN-2005 (first entry)
XX Murine cancer-associated genomic DNA #39.
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
XX cyostatic; gene; ds.
XX Mus sp.
XX WO2005031001-A2.
XX 07-APR-2005.
XX 23-SEP-2004; 2004WO-US031617.
XX 23-SEP-2003; 2003US-00669920.
XX (CHIR) CHIRON CORP.
XX Morris DW, Malandro MS;
XX WPI; 2005-273395/28.
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
XX comprises two or more nucleic acid probes.
XX Disclosure; SEQ ID NO 446; 198bp; English.

CC The invention relates to a nucleic acid array for detecting a cancer
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC The invention also relates to a peptide array comprising two or more
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC that binds to a polypeptide, an isolated antibody or its fragment which
CC binds to a polypeptide, which is prepared by immunizing a host animal
CC with a composition comprising the polypeptide or its antigen binding
CC fragment and collecting cells from the host expressing antibodies against
CC the antigen or its antigen binding fragment, a composition comprising the
CC antibody and a carrier, a method of screening for anticancer activity, a
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a
CC method of treating cancer and a method of inhibiting expression of a CA
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
CC nucleic acids. The antibody is useful for detecting the presence or
CC absence of cancer cells in an individual which involves contacting cells
CC from the individual with the antibody and detecting a complex of a CA
CC protein from the cancer cells and the antibody, where the detection of
CC the complex correlates with the presence of cancer cells in the
CC individual. The composition is useful for inhibiting growth of cancer
CC cells in an individual or for delivering a therapeutic agent to cancer
CC cells in an individual. The invention is also useful for diagnosing
CC cancer, for treating cancer and for inhibiting expression of a CA gene in
CC a cell. This sequence represents murine cancer-associated genomic DNA of
CC the invention.

XX Sequence 235962 BP; 68244 A; 47144 C; 47014 G; 72643 T; 0 U; 917 Other;
XX

Query Match 75.2%; Score 18.8; DB 14; Length 235962;
Best Local Similarity 90.9%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAAAGTCCAAATTCAG 22
AAAAAAAAAGTCCAAATTCAG 224065

Db 224044 AAAAAAAAAAGTCCAAATTCAG 224065

RESULT 48
AAV75702
ID AAV75702 standard; DNA; 102 BP.
XX AAV75702;
XX 16-MAR-1999 (first entry)
XX Staphylococcus aureus contig SEQ ID #1391.
XX
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
XX Staphylococcus aureus.
XX BP786519-A2.
XX 30-JUL-1997.
XX 07-JAN-1997; 97BP-00100117.
XX 05-JAN-1996; 96US-0009861P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Barash SC, Dillon PJ, Fannon MR, Rosen CA;
XX WPI; 1997-374922/35.
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -
XX stored on computer readable medium and used in the production of anti-
XX S.aureus vaccines.
XX Claim 1; Page 1942; 3271bp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the computer
CC readable medium

XX Sequence 102 BP; 31 A; 21 C; 17 G; 32 T; 0 U; 1 Other;
XX

Query Match 74.4%; Score 18.6; DB 2; Length 102;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 AAAAAAAAAAGTCCAAATTCAGATA 25
AAAAAAAAAAATATTCATTCATATATA 56

Db 32 AAAAAAAAAATTCATTCATATATA 56

```

RESULT 49
ADK94078/c
ID ADK94078 standard; DNA; 121 BP.
XX
AC ADK94078;
XX
DT 06-MAY-2004 (first entry)
XX
DE Polynucleotide used to detect SNPs of the invention #3107.
XX
KW human; single nucleotide polymorphism; SNP; de.
XX
OS Homo sapiens.
XX
PN JF2003259875-A.
XX
PD 16-SEP-2003.
XX
PF 08-MAR-2002; 2002JP-00064373.
XX
PR 08-MAR-2002; 2002JP-00064373.
XX
PA (KAGA-) KAGAKU GIUTSU SHINKO JIGYODAN.
XX
DR WPI; 2004-093977/10.
XX
PT Novel polynucleotide useful for PCR amplification along with two DNA
PT fragment from another set of sequences, or for detecting single
PT nucleotide polymorphism in human gene.
XX
PS Claim 1; SEQ ID NO 3107; 2627bp; Japanese.
XX
CC The present invention relates to a polynucleotide isolated from a human
CC gene and is useful for detecting a single nucleotide polymorphism in a
CC human gene or for diagnosing of disease. The invention enables the
CC detection of a single nucleotide polymorphism in a human gene. The
CC present sequence represents a polynucleotide used to detect SNPs of the
CC invention.
XX
SQ Sequence 121 BP; 31 A; 23 C; 14 G; 52 T; 0 U; 1 Other;
Query Match 74.4%; Score 18.6; DB 12; Length 121;
Best Local Similarity 84.0%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 59 AAAAAAAAAAGTCCAAAGTAACATA 35

RESULT 50
ABX79682/c
ID ABX79682 standard; cDNA; 123 BP.
XX
AC ABX79682;
XX
DT 17-APR-2003 (first entry)
XX
DE EST polymorphic DNA repeat polynucleotide #7.
XX
KW EST; expressed sequence tag; ss; polymorphic repeat; tandem repeat;
KW polymorphic marker prediction of ubiquitous simple sequences; POMPOUS;
KW Rep-X; human; genetic disease; drug-treatment; Machado-Joseph;
KW Haw River syndrome; Huntington's disease; fragile-X syndrome;
KW Friedrich's ataxia; myotonic dystrophy; hyperandrogenemia;
KW spinal atrophy; bulbar atrophy; spinocerebellar ataxia.
XX
OS Homo sapiens.
XX
PN US6472154-B1.
XX
PD 29-OCT-2002.
XX

```

```

PF 31-DEC-1999; 99US-00475947.
XX
PR 31-DEC-1999; 99US-00475947.
XX
PA (TEXA ) UNIV TEXAS SYSTEM.
XX
PI Garner HR, Wren JD, Minna JD, Fondon JW;
XX
DR WPI; 2003-200818/20.
XX
PT Identifying a candidate polymorphic repeat within a coding sequence, for
PT understanding or treating genetic disease, comprises detecting tandem
PT repeats in a target coding sequence and scoring the repeats for
PT polymorphic probability.
XX
PS Example; Col 143; 588bp; English.
XX
CC The invention discloses a method for identifying a candidate polymorphic
CC repeat within a coding sequence (expressed sequence tag, EST), which
CC comprises detecting tandem repeats in a target coding sequence, scoring
CC the repeats for polymorphic probability and generating a dataset
CC correlating the repeats with polymorphic probability to identify a
CC candidate polymorphic repeat. The computational methods (polymorphic
CC marker prediction of ubiquitous simple sequences, POMPOUS, and Rep-X) are
CC useful for identifying and detecting candidate polymorphic repeats in
CC human genes, which can be used to understand, treat or eliminate genetic
CC diseases, predispositions or adverse drug-treatment reactions. Examples
CC of diseases linked to nucleotide repeats are Machado-Joseph, Haw River
CC syndrome, Huntington's disease, fragile-X syndrome, Friedrich's ataxia,
CC myotonic dystrophy, hyperandrogenemia, spinal and bulbar atrophy and
CC spinocerebellar ataxia. The sequences presented in ABX79676-ABX80022 are
CC the polymorphic repeats identified for a search of human ESTs
XX
SQ Sequence 123 BP; 42 A; 19 C; 24 G; 38 T; 0 U; 0 Other;
Query Match 74.4%; Score 18.6; DB 8; Length 123;
Best Local Similarity 84.0%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 78 AAAAAAATATATTCATTTCATATA 54

Search completed: December 14, 2005, 02:43:10
Job time : 209.2 secs

```



```

C 96 20.2 80.8 630 11 CR964117 mch4-10YN
97 20.2 80.8 638 3 BM162997 EST565520
98 20.2 80.8 651 3 BM160714 EST563237
99 20.2 80.8 655 3 BM249338 EST562938
100 20.2 80.8 691 6 CR469204 PI7F04.P1
101 20.2 80.8 693 10 CL588053 OB_Ba008
102 20.2 80.8 697 8 DN335744 LIB3523-0
103 20.2 80.8 717 11 CR046745 Forward s
104 20.2 80.8 719 11 CM091591 104_453_1
105 20.2 80.8 720 8 DN129690 1143860 M
106 20.2 80.8 735 9 BH489900 B06PN27TF
107 20.2 80.8 750 10 AG517780 Nus muscu
108 20.2 80.8 759 10 CG972699 MBEG507R
109 20.2 80.8 777 8 CX451695 JGI_XZG23
110 20.2 80.8 784 7 CN754154 ID0AA12A
111 20.2 80.8 859 9 AZ689744 ENTKF57TF
112 20.2 80.8 865 8 DR562604 WS02625.C
113 20.2 80.8 871 7 CN761457 ID0AAA30B
114 20.2 80.8 909 8 DR469484 WS00947.B
115 20.2 80.8 914 10 CM957127 TCB52.1.D
116 20.2 80.8 953 9 AZ691792 ENTKM52TF
117 20.2 80.8 979 8 DN575085 91973824
118 20.2 80.8 997 3 BI465187
119 20.2 80.8 1000 3 B08139
120 20.2 80.8 1003 9 BI2235
121 20.2 80.8 1030 9 BI2507
122 20.2 80.8 1113 7 CK027949 AGENCOURT
123 20.2 80.8 1128 9 B10311
124 20.2 80.8 1229 7 CV673032 RET78U_10
125 20.2 80.8 637 7 C0081192 GR_BA45B
126 19.8 79.2 123 1 A1461086 aa73906.Y
127 19.8 79.2 320 1 BB097408 BB097408
128 19.8 79.2 331 3 B627421 B627421
129 19.8 79.2 402 2 BB636505 BB636505
130 19.8 79.2 412 2 BB02707 B84602.Y
131 19.8 79.2 487 2 CC117211 NDU_94F21
132 19.8 79.2 514 9 A2606591 IM0428002
133 19.8 79.2 521 8 CX076986 UCR808_5
134 19.8 79.2 547 1 AJ814680 AJ814680
135 19.8 79.2 548 2 BR013982 BR013982
136 19.8 79.2 582 10 CW769663 OG_BBa008
137 19.8 79.2 584 10 CL626212 OR_BBa002
138 19.8 79.2 609 9 AZ000772 RPCI-23-3
139 19.8 79.2 610 9 BZ883828 CH240_230
140 19.8 79.2 619 9 A0581955 RPCI-11-4
141 19.8 79.2 625 10 CE505871 tigr-g8s-
142 19.8 79.2 631 10 CE633287 tigr-g8s-
143 19.8 79.2 644 10 CL837167 OR_CBA006
144 19.8 79.2 652 3 BM598516 17006876
145 19.8 79.2 656 3 A0510231 nbdb0095A
146 19.8 79.2 659 3 BP704125 BP704125
147 19.8 79.2 668 10 CL744582 OR_BBa008
148 19.8 79.2 674 9 A0510354 nbdb0095B
149 19.8 79.2 685 10 CE605698 tigr-g8s-
150 19.8 79.2 700 10 CL813631 OR_CBA002

```

ALIGNMENTS

```

RESULT 1
LOCUS CR560919/c 698 bp mRNA linear EST 19-JUN-2004
DEFINITION CR560919 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA002c02
3', mRNA sequence.
ACCESSION CR560919
VERSION CR560919.1 GI:50390996
KEYWORDS EST.
ORGANISM Xenopus tropicalis (western clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.

```

```

REFERENCE
AUTHORS Croning,M.D.R., Ashurest,J.L., Taylor,R., Garrett,N. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: THDA002c02.g1kT7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: T7.
Location/Qualifiers
1..698
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA002c02"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Bacterichia coli DH10B."
/clone_lib="XGC-tailbud-head"
/notes="Vector: PCS107; Site_1: EcoRI; Site_2: NotI; CDNA
was oligo dt primed from sug of poly A+ RNA from tailbud
head. EcoRI-NotI cut CDNA was then ligated into PCS107
with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 89.6%; Score 22.4; DB 7; Length 698;
Best Local Similarity 95.8%; Pred. No. 6; g+e+2;
Matches 23; Conservative 1; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAAGTCCATTGAGAT 24
Db 277 AAAAAAAAAAGTTCAATTCAGAT 254

RESULT 2
LOCUS CR560918 710 bp mRNA linear EST 19-JUN-2004
DEFINITION CR560918 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA002c02
5', mRNA sequence.
ACCESSION CR560918
VERSION CR560918.1 GI:50390995
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE Croning,M.D.R., Ashurest,J.L., Taylor,R., Garrett,N. and Rogers,J.
AUTHORS Sanger Xenopus tropicalis EST project 2001 (2004)
TITLE Unpublished (2004)
JOURNAL Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: THDA002c02.p1kbsP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: SP6.
Location/Qualifiers
1..710
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA002c02"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Bacterichia coli DH10B."
/clone_lib="XGC-tailbud-head"
/notes="Vector: PCS107; Site_1: EcoRI; Site_2: NotI; CDNA

```

was oligo dT primed from 5ug of poly A+ RNA from tailbud head. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 89.6%; Score 22.4; DB 7; Length 710;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 461 AAAAAAAAAAGTTCATTCAGAT 484

RESULT 3
CR413014 711 bp mRNA linear EST 13-JUN-2004
LOCUS CR413014 XGC-tailbud Xenopus tropicalis cDNA clone TTBA063b04 5',
DEFINITION mRNA sequence.

ACCESSION CR413014.1 GI:48681261
VERSION CR413014
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 711)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk

FEATURES
source
1..711
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTBA063b04"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from 5ug of poly A+ RNA from tailbud. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
Seq primer: SP6.
Location/Qualifiers

FEATURES
source
1..711
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTBA063b04"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from 5ug of poly A+ RNA from tailbud. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
Seq primer: SP6.
Location/Qualifiers

Query Match 89.6%; Score 22.4; DB 7; Length 711;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 572 AAAAAAAAAAGTTCATTCAGAT 595

RESULT 4
CR565819 721 bp mRNA linear EST 19-JUN-2004
LOCUS CR565819 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA013e08
DEFINITION 3', mRNA sequence.
ACCESSION CR565819
VERSION CR565819.1 GI:50395896
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)

FEATURES
source
1..721
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA013e08"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud-head"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from 5ug of poly A+ RNA from tailbud head. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
Seq primer: T7.
Location/Qualifiers

ORGANISM

Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 721)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk

FEATURES
source
1..721
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA013e08"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud-head"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from 5ug of poly A+ RNA from tailbud head. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
Seq primer: T7.
Location/Qualifiers

ORIGIN

Query Match 89.6%; Score 22.4; DB 7; Length 721;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 276 AAAAAAAAAAGTTCATTCAGAT 253

RESULT 5
CX359920 757 bp mRNA linear EST 05-JAN-2005
LOCUS CX359920
DEFINITION JGI XZT958.Fwd NIH XGC tropTad5 Xenopus tropicalis cDNA clone IMAGE:7581097 5', mRNA sequence.
ACCESSION CX359920
VERSION CX359920.1 GI:57128479
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 757)
Richardson,P., Lucas,S., Rokhsar,D., Dettler,J.C., Ng,D.C., Brokstein,P. and Lindquist,E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project Unpublished (2004)
Other_ESTs: JGI XZT958.rev
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicalis.berkeley.edu/home>
cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
Clone Distribution: I.M.A.G.E. Consortium/BLND: <http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert. Plate: XZT 0009 row: 1 column: 23 High quality sequence stop: 730.

FEATURES

source

```

1..757
Location/Qualifiers
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7581097"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electrotan-Blue"
/clone_lib="NIH_XGC_tropTad5"
/notes="Vector: PCS108; Site_1: SalI; Site_2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector PCS108 (http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)"

```

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 757;
 Best Local Similarity 95.8%; Pred. No. 6.9e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
 |||
 Db 391 AAAAAAAAAAGTTCATTCAGAT 414

RESULT 6 CX387336 778 bp mRNA linear EST 05-JAN-2005

LOCUS

DEFINITION JGI_XZT21463.rev NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone
 IMAGE:7598032 3', mRNA sequence.

ACCESSION

VERSION CX387336
 KEYWORDS CX387336.1 GI:57155893

SOURCE

ORGANISM

EST.
 Xenopus tropicalis (western clawed frog)

Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus; Silurana.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

DOI Joint Genome Institute Xenopus tropicalis EST project
 Unpublished (2004)
 Other ESTs: JGI_XZT21463.fwd
 Contact: Lindquist, E.A., Richardson, P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 296 5600
 Fax: 925 296 5710
 Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicalis.berkeley.edu/home>
 cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
 DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
 Clone Distribution: I.M.A.G.E. Consortium/BLNI:
<http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
 Poly-A: Based upon the presence of a run of 14 or more T residues

at the beginning of the sequence, this clone was polyadenylated. The resulting Poly-T sequence has been removed. Plate: XZT 0221 row: n column: 14 High quality sequence stop: 693 POLY=A=Yes.

FEATURES

source

```

1..778
Location/Qualifiers
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7598032"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electrotan-Blue"
/clone_lib="NIH_XGC_tropTad5"
/notes="Vector: PCS108; Site_1: SalI; Site_2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector PCS108 (http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)"

```

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 778;
 Best Local Similarity 95.8%; Pred. No. 6.9e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
 |||
 Db 250 AAAAAAAAAAGTTCATTCAGAT 227

LOCUS

DEFINITION JGI_XZT38998.fwd NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone
 IMAGE:7615232 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS CX391469.1 GI:57172149

SOURCE

ORGANISM

EST.
 Xenopus tropicalis (western clawed frog)

Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus; Silurana.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

DOI Joint Genome Institute Xenopus tropicalis EST project
 Unpublished (2004)
 Other ESTs: JGI_XZT38998.rev
 Contact: Lindquist, E.A., Richardson, P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 296 5600
 Fax: 925 296 5710
 Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicalis.berkeley.edu/home>
 cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
 DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
 Clone Distribution: I.M.A.G.E. Consortium/BLNI:
<http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
 Plate: XZT 0405 row: k column: 6
 High quality sequence stop: 790.

FEATURES

Location/Qualifiers
1.793

/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7615232"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="B. coli XL1-Blue derivative, Stratagene Electrogen-Blue"
/clone_id="NIH_XGC_tropTad5"
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 793;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTCAGAT 24
|||||
Db 679 AAAAAAAAAAGTTCATTCAGAT 702

RESULT 8

CX329053 802 bp mRNA linear EST 04-JAN-2005
DEFINITION JGI_XZT67960.rev NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone

IMAGE:7787016 3', mRNA sequence.

CX329053
CX329053.1 GI:57065525

EST.

Xenopus tropicalis (western clawed frog)
Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 802)

Richardson, P., Lucas, S., Rohrsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)

Other ESTs: JGI_XZT67960.fwd

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home

cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/ILNL:

http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of
the JGI Clone ID and the direction of sequencing. The suffix '.rev'

indicates a reverse sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.

The resulting Poly-A sequence has been removed.

Small Insert: Based upon one or more sequencing reads of this clone
where vector sequence was present at both ends, this clone has been
determined to contain a cDNA insert on the order of 600-1000 bases.
Plate: XZT 0705 row: p column: 22

FEATURES

High quality sequence stop: 716
POLYA=Yes.
Location/Qualifiers
1..802

/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7787016"
/tissue_type="whole embryo"
/dev_stage="Tadpole (st. 36-41)"
/lab_host="B. coli XL1-Blue derivative, Stratagene Electrogen-Blue"
/clone_id="NIH_XGC_tropTad5"
/note="Vector: pCS108; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector pCS108
(http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 802;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTCAGAT 24
|||||
Db 250 AAAAAAAAAAGTTCATTCAGAT 227

RESULT 9

CX362328 813 bp mRNA linear EST 05-JAN-2005
DEFINITION JGI_XZT42336.fwd NIH_XGC_tropTad5 Xenopus tropicalis cDNA clone

IMAGE:7618033 5', mRNA sequence.

CX362328
CX362328.1 GI:57130887

EST.

Xenopus tropicalis (western clawed frog)
Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 813)

Richardson, P., Lucas, S., Rohrsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)

Other ESTs: JGI_XZT42336.rev

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of
California, Berkeley: http://tropicalis.berkeley.edu/home

cDNA Library Preparation: Richard M. Harland Laboratory, University
of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov

Clone Distribution: I.M.A.G.E. Consortium/ILNL:

http://image.llnl.gov

Naming Conventions: EST name is generated by the concatenation of
the JGI Clone ID and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.
Plate: XZT 0441 row: o column: 23
High quality sequence stop: 771.
Location/Qualifiers
1..813
/organism="Xenopus tropicalis"

FEATURES

source

/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7618033"
/issue_type="whole embryo"
/dev_stage="Tadpole (st.36-41)"
/lab_host="E. coli XL1-Blue derivative, Stratagene Electrogen-Blue"
/clone_id="NH_XGC_tropTad5"
/note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector PCS108
(http://mob.berkeley.edu/labs/harland/pages/plasmids.html)"

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 813;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTTCAGAT 24
|||||
Db 206 AAAAAAAAAAGTTCATTTCAGAT 229

RESULT 10
CR569461 815 bp mRNA linear EST 21-JUN-2004
DEFINITION CR569461 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA030103
5', mRNA sequence.
ACCESSION CR569461
VERSION CR569461.1 GI:50456887
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 815)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
Unpublished (2004)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: THDA030103.plkasp6
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.
Seq primer: SP6.
Location/Qualifiers
1..815
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA030103"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_id="XGC-tailbud-head"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA head. EcoRI-NotI cut cDNA was then ligated into pcs107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 89.6%; Score 22.4; DB 7; Length 815;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTTCAGAT 24

|||||
Db 765 AAAAAAAAAAGTTCATTTCAGAT 788

RESULT 11
CR413015/c 821 bp mRNA linear EST 13-JUN-2004
DEFINITION CR413015 XGC-tailbud Xenopus tropicalis cDNA clone TTDA063b04.3', mRNA sequence.
ACCESSION CR413015
VERSION CR413015.1 GI:48681262
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 821)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
Unpublished (2004)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TTDA063b04.q1k7
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.
Seq primer: T7.
Location/Qualifiers
1..821
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTDA063b04"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_id="XGC-tailbud"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from tailbud. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 89.6%; Score 22.4; DB 7; Length 821;
Best Local Similarity 95.8%; Pred. No. 6.9e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTTCAGAT 24
|||||
Db 269 AAAAAAAAAAGTTCATTTCAGAT 246

RESULT 12
CX329054 824 bp mRNA linear EST 04-JAN-2005
LOCUS CX329054
DEFINITION UGI XT67960.fwd NIH XGC tropTad5 Xenopus tropicalis cDNA clone IMAGE:7787016 5', mRNA sequence.
ACCESSION CX329054
VERSION CX329054.1 GI:57065526
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 824)
Richardson,P., Lucas,S., Rokhsar,D., Dettler,D.C., Ng,D.C., Brokstein,P. and Lindquist,E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other_Ests: UGI_XT67960.rev

Contact: Lindquist, E.A., Richardson, P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 296 5600
 Fax: 925 296 5710

Email: cdna@jgi-psf.org
 Tissue Procurement: Richard M. Harland Laboratory, University of California, Berkeley: <http://tropicalis.berkeley.edu/home>
 cDNA Library Preparation: Richard M. Harland Laboratory, University of California, Berkeley
 DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>
 Clone Distribution: I.M.A.G.E. Consortium/LNI: <http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of the UGI Clone id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
 Small Insert: Based upon one or more sequencing reads of this clone where vector sequence was present at both ends, this clone has been determined to contain a cDNA insert on the order of 600-1000 bases.
 Plate: X27 0705 row: P column: 22
 High quality sequence stop: 801.

FEATURES

source

```
1..824
  /organism="Xenopus tropicalis"
  /mol_type="mRNA"
  /db_xref="taxon:8364"
  /clone="IMAGE:7787016"
  /tissue_type="whole embryo"
  /dev_stage="Tadpole (st.36-41)"
  /lab_host="E. coli XL1-Blue derivative, Stratagene Electropen-Blue"
  /clone_id="NIH XGC tropTad5"
  /note="Vector: PCS108; Site 1: SalI; Site 2: NotI; Tadpole library constructed by Russell B. Fletcher in R. Harland's lab using poly A RNA and oligo dt primers (Invitrogen SuperScript Plasmid System for cDNA Synthesis and Cloning). SalI (5' end) -NotI (3' end) cDNA was inserted into vector PCS108
  (http://mcb.berkeley.edu/labs/harland/pages/plasmids.html)"
```

ORIGIN

Query Match 89.6%; Score 22.4; DB 8; Length 824;
 Best Local Similarity 95.8%; Pred. No. 6.9e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCAGATTCAGAT 24
 Db 623 AAAAAAAAAAGTTCAGATTCAGAT 646

RESULT 13
 LOCUS A1133803 426 bp mRNA linear EST 15-SEP-1998
 DEFINITION BSBMFS244K13K Brugia malayi microfilaria cDNA (SAM94LS-BmMf)
 ACCESSION A1133803
 VERSION A1133803.1 GI:3608104
 KEYWORDS EST.
 SOURCE Brugia malayi
 ORGANISM Brugia malayi
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Brugia.

REFERENCE 1 (bases 1 to 426)
 AUTHORS Williams, S.A.
 TITLE Genes expressed in microfilaria of Brugia malayi
 JOURNAL Unpublished (1995)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA

Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 EST submitted by Molecular Parasitology Group, New England Biolabs, Inc., 32 Tozer Road, Beverly, MA, 01915, USA. Email: dnaseq@neb.com

Method 2'. The numerical designation after the 52 refers to the microfilter tray number (01 - 48) followed by the letter and number of the microfilter tray position. Method '2' refers to sequenced clones which have been selected after hybridization subtraction using highly redundant clones present in the gridded library (18,000 mass-excised colonies gridded as a high density array on nylon filters). Colonies not represented in the probe sets were used as templates for the sequencing reactions.
 Seq primer: pBluescript SK.

FEATURES

source

```
1..426
  /organism="Brugia malayi"
  /mol_type="mRNA"
  /strain="TFS Labs"
  /db_xref="taxon:6279"
  /clone="BSBMFS244K13"
  /lab_host="XL1-Blue MR"
  /clone_id="Brugia malayi microfilaria cDNA (SAM94LS-BmMf)"
  /note="Vector: lambda UniZap XR; Site 1: EcoR I; Site 2: Xho I; Lymphatic filarial nematode parasite of humans. mRNA was prepared from microfilariae of Brugia malayi isolated from jirds and converted to double stranded cDNA using reverse transcriptase and oligo(dt) followed by RNase H and DNapol I. The library had 3.5 x 105 independent recombinants and average insert size was 900 base pairs. The library was constructed by Lori Saunders. The library is available from Dr. S.A. Williams, email genome@smith.edu."
```

ORIGIN

Query Match 87.2%; Score 21.8; DB 1; Length 426;
 Best Local Similarity 92.0%; Pred. No. 1.1e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCAGATTCAGATA 25
 Db 349 AAAAAAAAAAGTTCAGATTCAGAAA 373

RESULT 14
 LOCUS CG991942 487 bp DNA linear GSS 15-DEC-2003
 DEFINITION CH240_151005.TV CHORI-240 Bos taurus genomic clone CH240_151005, genomic survey sequence.
 ACCESSION CG991942
 VERSION CG991942.1 GI:39917721
 KEYWORDS GSS.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 487)
 AUTHORS Costa, J.N., Mota, M., and Caetano, A.R.
 TITLE Brazil's Contribution to End-Sequencing the Bovine BAC Library
 JOURNAL CHORI-240
 COMMENT Unpublished (2003)
 Other_GSSs: CH240_151005.TV
 Contact: Caetano AR
 Department of Biotechnology
 Embrapa Recursos Geneticos e Biotecnologia
 Parque Estacao Biologica, Pinal Av. W/5 Norte, Brasilia-DF C.P. 02372, 70770-900 Brasil
 Tel: 55 61 448 4778
 Fax: 55 61 340 3658
 Email: acaetano@cenargen.embrapa.br

Clones are derived from the bovine BAC library CHORI-240
(<http://www.chori.org/bacpac/bovine240.htm>).

Bases shown have phred quality value equal to or higher than 20.

Bases with quality value below 20 were masked with 'N'.
For BAC library availability, please contact Pieter de Jong

(pdjong@mail.choi.org).

Clones may be purchased from BACPAC Resources

(<http://www.chori.org/bacpac/ordering/information.htm>).

This work was undertaken as part of the International Bovine BAC

Mapping Consortium (IBMC) by Embrapa Recursos Genéticos e

Biotechnology with financing from Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq), Brazil

Plate: 151 row: 0 column: 05

Seq primer: T7

Class: BAC ends

High quality sequence stop: 487.

FEATURES
Location/Qualifiers

1..487

/organism="Bos taurus"

/mol_type="genomic DNA"

/strain="Breed: Hereford"

/db_xref="taxon:9913"

/clone="CH240_151005"

/sex="male"

/cell_type="Blood"

/clone_id="CHORI-240"

/note="Vector: pTARBA1.3; Site 1: MboI; Site 2: MboI;
Hereford bull 11 Dommo 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN

Query Match

Best Local Similarity 87.2%; Score 21.8; DB 10; Length 487;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATTCATTCAGATA 25

Db 339 AAAAAAAAAAGTCTCATTCGATA 363

RESULT 15

BM165368 504 bp mRNA linear EST 04-DEC-2001

LOCUS BM165368/c

DEFINITION EST567891 PYBS Plasmodium yoelii yoelii cDNA clone pYCM189 5' end.

ACCESSION BM165368

VERSION BM165368.1 GI:17311049

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii

ORGANISM Plasmodium yoelii yoelii

REFERENCE 1 (bases 1 to 504)

AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdy,A.B.,

Fraser,C.M. and Carnucci,D.J.

Plasmodium yoelii EST project at TIGR

Unpublished (2001)

CONTACT: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference

Reagent Resource Center, ATCC

<http://www.malaria.mr4.org/mr4pages/index.html>

Seq primer: ADP.

FEATURES
Location/Qualifiers

1..504

/organism="Plasmodium yoelii yoelii"

/mol_type="mRNA"

/strain="17XL"

/sub_species="yoelii"

/db_xref="taxon:73239"

/clone="PYCM189"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/clone_id="PYBS"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cBYJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridZAP vector and plasmid DNA

isolated."

ORIGIN

Query Match 87.2%; Score 21.8; DB 3; Length 504;

Best Local Similarity 92.0%; Pred. No. 1.1e+03;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAATTCATTCAGATA 25

Db 504 AAAAAAAAAAGTCTCATTCAGATA 480

RESULT 16

BM167726 511 bp mRNA linear EST 04-DEC-2001

LOCUS BM167726/c

DEFINITION EST570249 PYBS Plasmodium yoelii yoelii cDNA clone pYCP68 5' end.

ACCESSION BM167726

VERSION BM167726.1 GI:17300958

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii

ORGANISM Plasmodium yoelii yoelii

REFERENCE 1 (bases 1 to 511)

AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdy,A.B.,

Fraser,C.M. and Carnucci,D.J.

Plasmodium yoelii EST project at TIGR

Unpublished (2001)

CONTACT: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference

Reagent Resource Center, ATCC

<http://www.malaria.mr4.org/mr4pages/index.html>

Seq primer: ADP.

FEATURES
Location/Qualifiers

1..511

/organism="Plasmodium yoelii yoelii"

/mol_type="mRNA"

/strain="17XL"

/sub_species="yoelii"

/db_xref="taxon:73239"

/clone="PYCP68"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/clone_id="PYBS"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cBYJ mice infected with Py17XL

parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dt)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZap vector and plasmid DNA isolated."

ORIGIN

Query Match 87.2%; Score 21.8; DB 3; Length 511;
Best Local Similarity 92.0%; Pred. No. 1.1e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTCAGATA 25
|||||
504 AAAAAAAAAAGTCCATTTAAATA 480

Db 504 AAAAAAAAAAGTCCATTTAAATA 480

RESULT 17
C2407435/c
LOCUS C2407435
DEFINITION 1003118 RP42 Bos taurus genomic clone RP42-164M07, genomic survey
sequence.
ACCESSION C2407435
VERSION C2407435.1 GI:62294268
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 735)
Somtegard,T.S., Van Tassel,C.P., Capuco,A.V., de Jong,P., Matukumalli,L.K., Shade,L.S., Bosak,S., Rubenfield,M. and Gaabare,L.C.

TITLE Bovine BAC End Sequences from Library RPCI-42
JOURNAL Unpublished (2005)
COMMENT Contact: Tad S. Somtegard
Bovine Functional Genomics Laboratory
Animal and Natural Resources Institute
Bldg. 200 Rm2A BARC-East, Beltsville, MD 20705, USA
Tel: 3015048416
Fax: 3015048414
Email: tads@arl.barc.usda.gov

FEATURES

Clones are derived from the Bovine BAC library RPCI-42 (http://bacpac.chori.org/mbvline2.htm). For BAC library availability, please contact Pieter de Jong (pdejong@gmail.com). This work was undertaken as part of the international Bovine BAC Mapping Consortium (IBBMC) by USDA-ARS-BRGL.
Plate: 164 row: M column: 07
Seq primer: TAATACGACTCATATGAGG
Class: BAC ends.
Location/Qualifiers
1..735
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="RP42-164M07"
/sex="male"
/cell_type="Blood"
/clone_lib="RP42"
/note="Vector: pBACe3.6, Site_1: MboI; Site_2: MboI;

ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 735;
Best Local Similarity 92.0%; Pred. No. 1.1e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCATTCAGATA 25
|||||
265 AAAAAAAAAAGTCCATTCAGTTA 241

Db 265 AAAAAAAAAAGTCCATTCAGTTA 241

RESULT 18
AO870656
LOCUS AO870656
DEFINITION nbe0040017r CUCI Rice BAC library (EcoRI) Oryza sativa (japonica cultivar-group) genomic clone nbe0040017r, genomic survey
sequence.
ACCESSION AO870656
VERSION AO870656
KEYWORDS AO870656.1 GI:6221107
GSS.

SOURCE

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euhartioideae; Oryzeae; Oryza.
1 (bases 1 to 776)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Wing,R.A. and Dean,R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: GGAACAGTATGACATG
Class: BAC ends
High quality sequence start: 39
High quality sequence stop: 519.
Location/Qualifiers
1..776
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="japonica"
/cultivar="japonica"
/db_xref="taxon:39947"
/clone="nbe0040017r"
/issue_type="Leaf"
/lab_host="E. coli DH10B"
/clone_lib="CUCI Rice BAC library (EcoRI)"
/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University

ORIGIN BAC/EST Resource Center (www.genome.clemson.edu)."

Query Match 87.2% Score 21.8; DB 9; Length 776;
 Best Local Similarity 92.0%; Pred. No. 1.1e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTGAGATA 25
 80 AAAAAAAAAAGTTCATTGAGATA 104

RESULT 19
 AGS57295/c 761 bp DNA linear GSS 23-DEC-2004
 LOCUS Mus musculus molossinus DNA, clone:MSMg01-475H15.TJ, genomic survey
 DEFINITION
 ACCESSION AGS57295 GI:48317993
 VERSION AGS57295.1
 KEYWORDS Mus musculus molossinus (Japanese wild mouse)
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriaki, K. and Shiroyoshi, T.

TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis

JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
 PUBMED 15574823
 AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 TITLE Direct Submission

JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Shikoro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: hattori@gsc.riken.jp URL: http://ngp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp). The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp

PRIMERS
 Sequencing : TJ
 LIBRARY
 Vector : pBACe3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI

FEATURES
 source Location/Qualifiers
 1..761
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMg01-475H15.TJ"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 85.6% Score 21.4; DB 10; Length 761;
 Best Local Similarity 95.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTGAGA 23
 |||||

DB 524 AAAAAAAAAAGTTCATTGAGA 502

RESULT 20
 CR445028/c 823 bp mRNA linear EST 19-JUN-2004
 LOCUS CR445028 XGC-tailbud Xenopus tropicalis cDNA clone TTBA065p17.3', mRNA sequence.

ACCESSION CR445028
 VERSION CR445028.1 GI:48970615

KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM
 Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE
 AUTHORS Cronin, M.D.R., Ashurek, J.L., Taylor, R., Garrett, N. and Rogers, J.
 TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
 JOURNAL Unpublished (2004)
 COMMENT Contact: Croning MDR
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE_ID: TTBA065p17.q1kT7
 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.
 Seq primer: T7

FEATURES
 source Location/Qualifiers

1..823
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TTBA065p17"
 /dev_stage="tailbud (stage 28-30)"
 /lab_host="Baccharichia coli DH10B"
 /clone_lib="XGC-tailbud"
 /note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from tailbud. EcoRI-NotI cut cDNA was then ligated into PCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 85.6% Score 21.4; DB 7; Length 823;
 Best Local Similarity 95.7%; Pred. No. 1.6e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTTCATTGAGAT 24
 270 AAAAAAAAAAGTTCATTGAGAT 248

RESULT 21
 CR445027 844 bp mRNA linear EST 19-JUN-2004
 LOCUS CR445027 XGC-tailbud Xenopus tropicalis cDNA clone TTBA065p17.5', mRNA sequence.

ACCESSION CR445027
 VERSION CR445027.1 GI:48970614

KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM
 Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.

REFERENCE
 AUTHORS Cronin, M.D.R., Ashurek, J.L., Taylor, R., Garrett, N. and Rogers, J.
 TITLE Sanger Xenopus tropicalis EST project 2001 (2004)
 JOURNAL Unpublished (2004)
 COMMENT Contact: Croning MDR
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPCALIS_SEQUENCE_ID: TTBA065p17.plksp6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: SP6.

FEATURES

source

Location/Qualifiers
1..844
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTBA065p17"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_id="XGC-tailbud"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5' end of poly A+ RNA from tailbud.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 85.6%; Score 21.4; DB 7; Length 844;
Best Local Similarity 95.7%; Pred. No. 1.6e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAGTTCATTCAGAT 24
|||||
Db 807 AAAAAAAAAAGTTCATTCAGAT 829

RESULT 22

LOCUS

CB254968 487 bp mRNA linear EST 31-DEC-2004
54-B010931-019-003-K13-T7R MP1Z-ADIS-019 Arabidopsis thaliana cDNA
clone MP1ZP768K13Q 5-PRIME, mRNA sequence.

ACCESSION

CB254968

VERSION

CB254968.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Arabidopsis thaliana (chale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 487)
Jakoby,M., Stracke,R., Soerensen,T.R. and Weishaar,B.
Arabidopsis thaliana cDNA library enriched in transcription factors
Unpublished (2003)
Contact: Weishaar B
ADIS DNA core facility at MP1Z
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weishaar@mpiz-koeln.mpg.de
Insert Length: 487 Std Error: 0.00
Plate: 3 row: K column: 13
Seq primer: T7R; CTATACGACTCATATAGGA.

FEATURES

source

Location/Qualifiers
1..487
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cuiivar="At7"
/db_xref="GABI:597426"
/db_xref="taxon:3702"
/clone="MP1ZP768K13Q"
/tissue_type="hypocotyl"
/dev_stage="tissue culture"
/lab_host="E. coli DH5alpha mcr"
/clone_id="MP1Z-ADIS-019"
/note="Vector: pSPORT1; Site 1: NotI primer adapter;
Site 2: SalI primer adapter; RNA from cellculture (At7) 5
days after inoculation treated with 0.002 mM cycloheximide
for 2 h in the dark. Sequence submission managed by
RZPD/GABI-Primary database: <http://gabi.rzpd.de>. This

clone is available from RZPD; contact RZPD (clone@rzpd.de)
for further information."

ORIGIN

Query Match 84.0%; Score 21; DB 6; Length 487;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAGTTCATTCAG 22
|||||
Db 456 AAAAAAAAAAGTTCATTCAG 436

RESULT 23

LOCUS

B2917317/c 549 bp DNA linear GSS 26-NOV-2004
CH240_102P10_TV CHORI-240 Bos taurus genomic clone CH240_102P10,
genomic survey sequence.

ACCESSION

B2917317

VERSION

B2917317.2

KEYWORDS

SOURCE

ORGANISM

REFERENCE
AUTHORS

1 (bases 1 to 549)
Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.A.,
Bachman,S., Green,C., Wright,C.L., Campos,E.J., Benson,L.D.,
Edwards,J., Liu,L., Osogawa,K., Womack,J.E., de Jong,P.J. and
Lewin,H.A.
A cattle-human comparative map built with cattle BAC-ends and human
genome sequence
Genome Res. 13 (8), 1966-1972 (2003)
12902387

TITLE

JOURNAL

PUBMED

COMMENT

On Nov 26, 2004 this sequence version replaced gi:31642703.
Other GSSs: CH240_102P10.TJ
Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Clones are derived from the bovine BAC library CHORI-240
(<http://www.chori.org/bacpac/bovine240.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(<http://www.chori.org/bacpac/ordering/information.htm>). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by the University of Illinois at Urbana
Champaign, USA with funds provided by grant No. AG202-34480-11828
from USDA-CRRES and AG99-35205-8534 from USDA/NRI (Livestock
Genome Sequencing Initiative)
Plate: 102 row: P column: 10
Seq primer: T7
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..549
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="bred: Hereford"
/db_xref="taxon:9913"
/clone="CH240_102P10"
/sex="Male"
/cell_type="blood"
/clone_id="CHORI-240"
/note="Vector: pTARPA3.3; Site 1: MboI; Site 2: MboI;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 84.0%; Score 21; DB 9; Length 549;

Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTCA 21
Db 256 AAAAAAAAAAGTCCATTCA 236

RESULT 24

CN640041/c

LOCUS

DEFINITION CN640041 182 bp mRNA linear EST 30-SEP-2004

261A11_554559 Douglas-fir cDNA library PmIFG_73-6 Pseudotsuga

menziesii var. menziesii cDNA clone Df261-A11 5, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

COMMENT

Neale, D.B., Troggio, M., Brown, G.R., Jernstedt, K.D. and Krutovsky, K.V., Comparative mapping in the Pinaceae Genetics 168 (1), 447-461 (2004)
Contact: Krutovsky KV
David Neale's lab
Institute of Forest Genetics, USDA Forest Service, Pacific Southwest Research Station
Department of Plant Sciences, University of California, One Shields Avenue, Davis, CA 95616, USA
Tel.: 1-530-752-8412
Fax: 1-530-754-9366
Email: kkrutovsky@ucdavis.edu
PCR Primers
FORWARD: M13 forward at 5' end
Seq primer: M13 forward at 5' end
High quality sequence stop: 182.

FEATURES

Location/Qualifiers

1..182

/organism="Pseudotsuga menziesii var. menziesii"

/mol_type="mRNA"

/db_xref="taxon:278161"

/clone="DF261-A11"

/tissue_type="entire seedlings"

/dev_stage="one month old seedlings"

/note="Vector: pGEM-3Z; Site 1: EcoICRI; Site 2: EcoICRI; The cDNA library was obtained from one month old Douglas-fir seedlings. Total RNA was extracted from ~5 g of ground tissue that consisted of the mixture of 10 seedlings following the protocol described in Chang et al. (1993). mRNA was isolated from the total RNA using the PolyAtract mRNA Isolation System III (Promega Z5300). Double stranded cDNA was prepared using the Universal Ribocloner cDNA Synthesis System (Promega C4360) and filtered through a Sephadryl S-400 column to remove low molecular weight products of less than 300 bp. Double-stranded blunt-end cDNA was ligated into the EcoICRI-cut dephosphorylated pGEM-3Z sequencing vector and electroporated into E. coli DH5 alpha. cDNA clones were sequenced by Integrated Genomics, Inc. (Chicago, USA)."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 182;

Best Local Similarity 91.7%; Pred. No. 2.6e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTCA 24

Db 158 AAAAAAAAAAGTCAATTCA 135

RESULT 25

CL378278/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CL378278 308 bp DNA linear GSS 19-AUG-2004
RPC14_423B6.r RPC1-44 Sus scrofa genomic clone RPC14_423B6,
genomic survey sequence.
CL378278
CL378278.1 GI:51430243
GSS.
Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suidae;
Sus.
1 (bases 1 to 308)
Kogatcheva, M.B., Meyers, S., He, W., Larkin, D.M., Marron, B.M.,
Beever, V.E. and Schork, L.B.
PI99Y-BACing the Human Genome: Constructing a Porcine Physical Map
Through Comparative Genomics
Unpublished (2004)
Contact: Lawrence B. Schork
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 265 5326
Fax: 217 244 5617
Email: schork@uiuc.edu
Clones are derived from the porcine BAC library RPC1-44
(<http://www.bacpac.chori.org/porcine24.htm>). For BAC library
availability, please contact Pieter de Jong (pdjong@chori.org).
Clones may be purchased from BACPAC Resources
(<http://BACPACresources.chori.org>). This work was undertaken as part
of the International Swine Genome Sequencing Consortium by
University of Illinois at Urbana Champaign, USA with funds provided
by grant No. AG2002-34480-11828 from USDA-CSREBS and
AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
Initiative)
Plate: 423 row: B column: 6
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..308
/organism="Sus scrofa"
/mol_type="genomic DNA"
/strain="four pigs (breed: 37.5% Yorks Landrace and 25%
Meishan)"
/db_xref="taxon:9823"
/clone="RPC14_423B6"
/sex="male"
/cell_type="blood"
/clone_lib="RPC1-44"
/note="Vector: pTARBAC2; Site 1: EcoRI; Site 2: EcoRI;
porcine male BAC library produced by Pieter de Jong"

ORIGIN

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 10; Length 308;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTCA 24

Db 149 AAAAAAAAAAGTCAATTCA 126

RESULT 26

CR571128

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

CR571128 362 bp mRNA linear EST 21-JUN-2004
CR571128 XGC-tailbud-head Xenopus tropicalis cDNA clone THDA024106
3', mRNA sequence.

CR571128
CR571128.1 GI:50458554
EST.

SOURCE
ORGANISM Xenopus tropicalis (western clawed frog)

REFERENCE
AUTHORS Amphibia; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Anphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 362)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
JOURNAL Unpublished (2004)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: THDA024106.q1kr7
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Nigel Garrett.
Seq primer: 17.

FEATURES
source
1..362
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="THDA024106"
/dev_stage="tailbud head (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud-head"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from tailbud head. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 83.2%; Score 20.8; DB 7; Length 362;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCAAATTCAGAT 24
Db 219 AAAAAAAAAAGTCAATTCMAAT 242

RESULT 27
LOCUS CX638428/c 371 bp mRNA linear EST 18-JAN-2005
DEFINITION UCRPT02.34G12.g Poncinus trifoliata Roots with Iron Deficiency -
UCRPT02.34G12-M23-1-5.g, mRNA sequence.
ACCESSION CX638428
VERSION CX638428.1 GI:57873257
KEYWORDS EST.
SOURCE Poncinus trifoliata
ORGANISM Poncinus trifoliata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids 1; Sapindales; Rutaceae; Poncinus.
1 (bases 1 to 371)
Close,T.J., Roose,M.L., Parker,D.R., Federici,C.F., Mandal,J., Fenton,R.D., Wamamaker,S., Landry,B., Hubert,N., Laforest,M., Landry,J. and Ligonde, A.
Development of EST Resources and New Genetic Markers for California Citrus - Poncinus trifoliata Roots with Iron Deficiency -
JOURNAL UCRPT02-UCR1
COMMENT Unpublished (2005)
Contact: Timothy J. Close
Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timochy.close@ucr.edu
Seq primer: 73.
Location/Qualifiers

SOURCE
1..371
/organism="Poncinus trifoliata"
/mol_type="mRNA"
/cultivar="Pomeroy"
/db_xref="taxon:37690"
/clone="UCRPT02.34G12-M23-1-5.g"
/rissue_type="root"
/dev_stage="seedling"
/lab_host="E. coli JYCl21"
/clone_lib="Poncinus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1"
/note="Vector: lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plant materials were prepared by Federici (Roose lab) with advice from Parker. Seedlings 138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO2 buffering. This solution was replaced on July 6, 2004. Roots from three plants were sampled on June 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using TRIzol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were placed, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wamamaker (Close lab) using the Harvest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a pruned 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Wamamaker). Sequences that survived all removal steps were submitted to Genbank."

ORIGIN
Query Match 83.2%; Score 20.8; DB 8; Length 371;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTCCAAATTCAGAT 24
|||||||

Db 193 AAAAAAAAAATTCAATTCAGAT 170

RESULT 28
BG511457/c 384 bp mRNA linear EST 22-JUL-2004
LOCUS rad02b10.y1 Gm-c1073 glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION ID: Gm-c1073-1243 5', mRNA sequence.
ACCESSION BG511457
VERSION BG511457.1 GI:13482114
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 384)
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Expelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,D., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave, Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
High quality sequence stop: 372.
Location/Qualifiers
1..384
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1073-1243"
/tissue_type="seedlings induced for symptoms of SDS (Sudden Death Syndrome) disease"
/dev_stage="2-3 weeks old"
/lab_host="DH10B"
/clone_1fb="Gm-c1073"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from 2-3 week old seedlings that were induced for symptoms of SDS (Sudden Death Syndrome) disease by the translocation of culture filtrate of Fusarium solani f. sp. glycines (Plant Cell Report 18:375-380). Cultivar Williams 82 is susceptible to the disease SDS. Plant tissue (expanded leaves, folded leaves, and new shoots) were collected at 1, 6, 24, and 48 hrs. after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene pBluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. Plants were inoculated by Shuxian Li (Glen Hartman lab, University of Illinois). Library was constructed by Reena Philip and Steve Clough

ORIGIN (Lila Vodkin lab, University of Illinois)."

Query Match 83.2%; Score 20.8; DB 2; Length 384;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 2 AAAAAAAAAATTCAATTCAGATA 25
Db 160 AAAAAAAAAAGTTCAATTCAGATA 137

RESULT 29
BE821144 489 bp mRNA linear EST 24-MAY-2001
LOCUS GM700014A20A5 Gm-r1070 Glycine max cDNA clone Gm-r1070-5386 3',
DEFINITION mRNA sequence.
ACCESSION BE821144
VERSION BE821144.1 GI:10253378
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 489)
AUTHORS Vodkin,L., Kelm,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Expelding,J., Raph,C., Shoop,B., Pardinas,J., Liu,L. and Lewin,H.
TITLE A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL Unpublished (1999)
COMMENT Other ESTs: AM509350 corresponding to Gm-c1016-7856 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Ewin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomeystems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
Location/Qualifiers
1..489
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/db_xref="taxon:3847"
/clone="Gm-r1070-5386"
/clone_1fb="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected for development of soybean that various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
Rerecking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The

corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'.

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 489;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Yr 2 AAAAAAAAAAGTCCATTGAGATA 25
|||||
Db 157 AAAAAAAAAAGTGAATTGAGATA 180

RESULT 30

LOCUS BG793768 512 bp mRNA linear EST 16-MAY-2001
DEFINITION UTSW_SMI5H5 UTSW Adult Mouse Skeletal Muscle Library Mus musculus
CDNA clone UTSW_SMI5H5, mRNA sequence.

ACCESSION BG793768.1 GI:14129338
VERSION EST.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Eumetazoa; Eumetazoa; Eumetazoa; Eumetazoa;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Muridae; Mus.

REFERENCE 1 (bases 1 to 512)
AUTHORS Gallardo, T.D., Schegeman, J.J., Pertsemilidze, A., Garner, H.R.,
Williams, R.S., and Shohet, R.V.

TITLE UTSW Adult Mouse Skeletal Muscle cDNA
Library

JOURNAL Unpublished (2001)
COMMENT Contact: Schegeman JJ
Shohet/Garner Labs
University of Texas Southwestern Medical Center
6000 Harry Hines Blvd., N2A.226, Dallas, TX 75390, USA
Tel: 214 648 1674

Email: Jeff.Schegeman@UTSouthwestern.edu
cDNA library constructed by UTSW as a component of the Program for
Genomic Applications (PGA) and the Reynolds Heart Disease
Prevention Grants for use in cDNA microarray experiments. Sequence
Quality: Sequence ends were trimmed based on percentage of ambigu-
ous base calls or 'N's in windowed segments. Sequencing: First-pass
sequencing; ABI Prism 377 sequencer and analysis software.
Seq primer: M13/PUC Reverse.

FEATURES

SOURCE

1. 512
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="UTSW_SMI5H5"
/sex="Pooled"
/tissue_type="Diaphragm/Hind limb muscles"
/cell_type="Skeletal muscle"
/dev_stage="2 months"
/lab_host="DH5a"
/clone_id="UTSW Adult Mouse Skeletal Muscle Library"
/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Mannatris); Cloning Technique: CUA Cloning (Clontech).
Life Technologies; Average insert size: 1.8 kb;
Insertion site: TACCTCAGTGAATTCGAGTC--> Other
information regarding entire library may be found at
http://pga.swmed.edu/Data/libraries/microarray_cdna_librar-
ies.htm."

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 512;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Yr 2 AAAAAAAAAAGTCCATTGAGATA 25
|||||
Db 237 AAAAAAAAAAGTGAATTGAGATA 260

RESULT 31

AZ434046 553 bp DNA linear GSS 03-OCT-2000
LOCUS IM0220E18P Mouse 10kb plasmid UUGCM library Mus musculus genomic
DEFINITION clone UUGCM0220E18 F, genomic survey sequence.

ACCESSION AZ434046
VERSION A2434046.1 GI:10558059
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Muridae; Mus.

REFERENCE 1 (bases 1 to 553)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
Niederhausen, A., and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0220 row: B column: 18
Seq primer: CGTTGTAACGACGCCACGT
Class: plasmid ends
High quality sequence stop: 553.

FEATURES

SOURCE

1. 553
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCM0220E18"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_id="Mouse 10kb plasmid UUGCM library"
/note="Vector: pMD29ny. Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi14732114|gb|AF129072.1) a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 553;

Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTCAGAT 24
Db 325 AAAAAAAAAAGTCCATTCAGAT 302

RESULT 32
B1785075/c 554 bp mRNA linear EST 21-JUN-2004
LOCUS B1785075
DEFINITION B1785075. y3 Gm-c1079 glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-c1079-1818 5' similar to TR:Q9ZT88 Q9ZT88 HYPOTHETICAL 77.8
KD PROTEIN ; mRNA sequence.

ACCESSION B1785075
VERSION B1785075.1 GI:15812800
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 554)
Shoemaker, R., Kaim, P., Vodkin, L., Epeiding, U., Coryell, V.,
Khan, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wyllie, T., Underwood, K., Stepien, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCam, R., Waterston, R., and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: east@wustl.edu

When it has been determined, an EST from the other end of this
clone is listed in the 'Other ESTs on clone' field. This clone is
available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
Seq primer: -40RP from Gibco
High quality sequence stop: 421.

FEATURES
Source
Location/Qualifiers
1..554
/organism="Glycine max"
/mol_type="mRNA"
/cultiivar="Bragg"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1079-1818"
/cissue_type="Roots of 8 day old 'Bragg' seedlings"
/dev_stage="8 days old"
/lab_host="DH10B"
/clone_lib="Gm-c1079"
/note="Vector: pBluescript II SK+, Site 1: EcoRI; Site 2:
XhoI. The mRNA was isolated from roots of 8 day old
'Bragg' seedlings that were mock-infected 72 hours prior
to harvest. Dr. Gary Stacey generously donated the
tissue. The roots were flash-frozen in liquid nitrogen.
Stratagene's cDNA Synthesis Kit (catalog number 200401)
was used to synthesize the cDNA. First-strand synthesis
was performed with 5-methyl dCTP, hence the ligated cDNA
was hemimethylated. A modification of Stratagene's
first-strand synthesis primer was used. An 'anchor'
nucleotide (V=A, C, or G) was added to the 3' end of the
primer [GAGACAGAGAGAGAGAGACTGCTCAG(T)18V] to anchor
the primer at the 5' end of the poly(A) tract. After
second-strand synthesis, the cDNA ends were filled in with
cloned Pfu DNA, ligated to EcoRI adapters and subsequently
phosphorylated. The cDNA was then precipitated and
redissolved in sterile, RNase-, DNase-free water. The XhoI

site within the first-strand synthesis primer was then
restricted by digestion with XhoI from Promega (40U/ul);
all XhoI sites in the cDNA would be protected by their
hemimethylated status. The cDNA constructs were
size-fractionated with a 500bp cutoff, using Sephacryl
S-500 High Resolution (Pharmacia Biotech) in a 2-mm
diameter column and a bed volume of approximately 1ml. The
column eluent was precipitated, redissolved, and ligated
into Stratagene's pBluescript II XR Predigested vector
(pBluescript II SK(+)) vector that has been digested with
EcoRI and XhoI, and phosphorylated by Stratagene. This
library was constructed in the laboratory of Dr. Paul Kaim
and Dr. Virginia H. Coryell at Northern Arizona
University."

ORIGIN
Query Match 83.2%; Score 20.8; DB 3; Length 554;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCATTCAGATA 25
Db 548 AAAAAAAAAAGTCCATTCAGATA 525

RESULT 33
CR792524/c 558 bp DNA linear GSS 24-SEP-2004
LOCUS CR792524
DEFINITION CR792524.1 genomic survey sequence.
ACCESSION CR792524
VERSION CR792524.1 GI:52673366
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 558)
Eggen, A., Schibler, L., and Roy, A.
Bovine BAC End Sequences from the INRA bovine BAC library
Unpublished
2 (bases 1 to 558)
Genoscope.
Direct Submission
Submitted (20-SEP-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Contact: Andre Eggen
Department of Animal Genetics - LGBC
INRA
78350 Jouy-en-Josas, France
Tel: 33 1 34 65 24 24
Fax: 33 1 34 65 24 78
Email: eggen@jouy.inra.fr
Clones are derived from the INRA bovine BAC library
(http://locus.jouy.inra.fr/fpc/cattle_bac_map.htm). For BAC library
availability, please contact Andre Eggen (eggen@jouy.inra.fr). This
work was undertaken as part of the International Bovine BAC
Mapping Consortium (IBBMC) by INRA (Jouy-en-Josas) and Genoscope
(Evry) Plate: 28 row: G column: 10
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
Source
Location/Qualifiers
1..558
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="Breed: Holstein"
/db_xref="taxon:9913"
/clone="INRA_28G10"
/sex="Male"
/cell_type="fibroblast"
/clone_lib="INRA bovine BAC"

ORIGIN

/note="Vector: pBel0AC11, Site 1: HindIII, Holstein bull;
INRA Bovine BAC library (Male) produced by Andre Eggen
Genoscope sequence ID : GR05AALAC09RM2"

Query Match 83.2%; Score 20.8; DB 11; Length 558;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGAGAT 24
|||||
Db 334 AAAAAAAAAAGTCCATTGAT 311

RESULT 34

BUT61408 605 bp mRNA linear EST 02-JUL-2004
sa871609.y1 Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID:
Gm-cl036-9809.5' similar to TR:Q92788 Q92788 HYPOTHETICAL 77.8 KD
PROTEIN, mRNA sequence.

ACCESSION BUT61408
VERSION BUT61408.1 GI:23726811
KEYWORDS EST.
SOURCE Glycine max (soybean)

ORGANISM Glycine max (soybean)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE

1 (bases 1 to 605)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harey, N.,
Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCam, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

TITLE

Public Soybean EST Project

COMMENT

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this
clone is listed in the 'Other ESTs on clone' field. This clone is
available through: Biogenetic Services, 801 32nd Ave, Brookings, SD
57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
Seq primer: -408P from Glibco
High quality sequence stop: 450.

FEATURES

source

1..605
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Jack"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl036-9809"
/tissue_type="somatic embryos cultured on MSD 20"
/lab_host="DH10B"
/clone_lib="Gm-cl036"
/note="Vector: pSPORT1, Site_1: NotI, Site_2: SalI; This
cDNA library was constructed from mRNA isolated from
somatic embryos (age ranging from 2 months to 9 months)
cultured on MSD 20. The library was prepared using the
Life Technologies superscript cDNA library construction
kit. Complementary DNA was synthesized from mRNA using a
poly (dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli Electromax DH10B host cells.

ORIGIN

This library was constructed in the laboratory of Dr. Lila
Vodkin by Anu Khanna at the University of Illinois at
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

Query Match 83.2%; Score 20.8; DB 5; Length 605;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTCCATTGAGATA 25
|||||
Db 414 AAAAAAAAAAGTCCATTGAGATA 391

RESULT 35

AZ565184 608 bp DNA linear GSS 07-MAY-2001
LOCUS 205PVE07 PV MBN #16 (amplified twice) Plasmodium vivax genomic 3'
DEFINITION genomic survey sequence.

ACCESSION AZ565184
VERSION AZ565184.1 GI:13971020
KEYWORDS GSS.
SOURCE Plasmodium vivax (malaria parasite P. vivax)
ORGANISM Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 608)
Cartlon, J.M.-R. and Dame, J.B.
The Plasmodium vivax and P. berghel gene sequence tag projects
Parasitol. Today 16 (10), 409 (2000)

TITLE

Parasitol. Today 16 (10), 409 (2000)

JOURNAL

Parasitol. Today 16 (10), 409 (2000)

PUBMED

11006469

Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.

FEATURES

source

1..608
/organism="Plasmodium vivax"
/mol_type="genomic DNA"
/strain="Belem"
/db_xref="taxon:5855"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/clone_lib="PV MBN #16 (amplified twice)"
/note="Vector: Lambda ZAP II (Stratagene); individual
clones excised into phagemid plasmids; Site 1: EcoR I;
Site 2: EcoR I; Genomic DNA was prepared from asynchronous
blood stage forms of the Belem line of P. vivax grown in
squirrel monkeys. Parasitized erythrocytes were purified
from contaminating host leukocytes by filtration of ADP
activated blood through acid-washed glass beads and
Whatman CF11 cellulose columns by gravity filtration.
Purified DNA was digested with mung bean nuclease in the
presence of 42.5% formamide at 500C as described
(Galinsek, M. et al. 1992. Cell 69,1213-1226; Vernick,
K.D. et al. 1988. N.A.R. 16, 6883-6896). Eco RI linkers
were added and the constructs ligated into Lambda ZAP II.
P. vivax Belem was originally isolated from a patient in
Belem, Brazil 1980 by Mercia de Arruda, adapted to Saimiri
monkeys by Urig Gysin, and maintained since 1983 in
squirrel monkeys."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 608;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGAGAT 24
|||||

Db 424 AAAAAAAAAAGTTCCAAATAGAT 447

RESULT 36
CE209996/c 654 bp DNA linear GSS 25-SEP-2003
LOCUS tigr-gss-dog-17000372701962 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE209996
VERSION CE209996.1 GI:35365651
KEYWORDS
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.

REFERENCE
AUTHORS 1 (bases 1 to 654)
Kirkhness, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkhness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkhness@tigr.org
Class: shotgun.

FEATURES
source location/Qualifiers
1..654
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: Bactxi; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 83.2%; Score 20.8; DB 9; Length 654;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCCAAATTCAGAT 24
Db 350 AAAAAAAAAAGATCCAAATTCAGAT 327

RESULT 37
CR348763 666 bp DNA linear GSS 17-NOV-2004
LOCUS mtel-83G9FM1 BAC end, cultivar Jemalong A17 of Medicago truncatula,
DEFINITION genomic survey sequence.
ACCESSION CR348763
VERSION CR348763.1 GI:45121282
KEYWORDS
SOURCE Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.

REFERENCE
AUTHORS 1 (bases 1 to 666)
Genoscope.
Direct Submission
Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Location/Qualifiers
1..666

FEATURES
source

ORIGIN
Query Match 83.2%; Score 20.8; DB 11; Length 666;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTTCCAAATTCAGATA 25
Db 260 AAAAAAAAAAGATCCAAATTCAGATA 283

RESULT 38
CX073947 691 bp mRNA linear EST 14-DEC-2004
LOCUS UCRCS08-36B04.9 Parent Washington Navel Orange Callus cDNA library
DEFINITION UCRCS08-2 Citrus sinensis cDNA clone UCRCS08-36B04-D8-1-5-9, mRNA
sequence.
ACCESSION CX073947
VERSION CX073947.1 GI:56587937
KEYWORDS
SOURCE Citrus sinensis
ORGANISM Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE
AUTHORS 1 (bases 1 to 691)
Close, T.J., Rose, M.L., Ye, X.R., Fenton, R.D., Manamaker, S.,
Lyons, M., Dang, C., Quintillo, C., Ikeda, J., Collin, M., Kacur, Y.,
Landry, B., Hubert, N., Laforest, M., Landry, J., and Ligonde, A.,
Development of EST Resources and New Genetic Markers for California
Citrus - Washington Navel Orange Callus - UCRCS08-2
Unpublished (2004)
Contact: Timothy J. Close
Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timothy.close@ucr.edu
Seq primer: T3.

JOURNAL
COMMENT

FEATURES
source location/Qualifiers
1..691
/organism="Citrus sinensis"
/mol_type="mRNA"
/cultivar="Washington navel"
/db_xref="taxon:2711"
/clone="UCRCS08-36B04-D8-1-5-9"
/tissue_type="Callus"
/dev_stage="Embryogenic and embryoid"
/lab_host="E. coli TUC121"
/clone_lib="Parent Washington Navel Orange Callus cDNA
library UCRCS08-2"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Parent Washington navel
orange embryogenic callus was established from undeveloped
ovules of ca. 10 mm diameter young fruits under open
pollination on Murashige-Skoog medium at 25°C with 16 h
light in a tissue culture room. Embryogenic callus,
globular and heart stage embryoids were pooled in
approximately equal portions in RNAlater (Ambion), then
RNA was extracted using TRIzol Reagent (Invitrogen).
Poly(A) RNA was purified from 500 microgram of total RNA
using Oligen Oligotex. A primary cDNA library was produced
using a lambda ZAP XR cDNA Synthesis Kit (Stratagene).
These steps were performed by Xinrong Ye (Roose lib, UC

Riverside). One million pfu from the primary library were mass excised to produce a phagemid population by Raymond Fenton (Close lab, UC Riverside). Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde, (by Close), then processed at UC Riverside (by Wanmaker, (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Rose, Federici, Wanmaker, Lyon, Ye, Jang, Collin, Kacar, Ikeda, Quintilio). Sequences that survived all removal steps were submitted to Genbank."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 691;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1
112
AAAAAAAAAGTTCCAATTCAGT 24
AAAAAAAAAGTTCCAATTCAGT 135

RESULT 39
BH922369
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BH922369 698 bp DNA linear GSS 01-OCT-2002
cdh01e02.g1 B.oleracea02 Brassica oleracea genomic, genomic survey
sequence.
BH922369
GSS.
GI:23402412
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 698)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rabinowicz, P.D. and Wilson, R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submis@wustl.edu
Plate: cdh01 row: e column: 02
Seq primer: -28RPPOT reverse
Class: Shotgun
High quality sequence start: 17
High quality sequence stop: 543.
Location/Qualifiers

FEATURES

Source
1. .698
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lhb="B.oleracea002"
/note="Vector: POTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 698;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1
114
AAAAAAAAAGTTCCAATTCAGT 24
AAAAAAAAAGTTCCAATTCAGT 238

RESULT 40
AG013987/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AG013987 711 bp DNA linear GSS 16-FEB-2005
Homo sapiens genomic DNA, 21q region, clone: B335C24X38, genomic
survey sequence.
AG013987
GSS.
GI:55789002
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1
Hattori, M., Ishii, K., Toyoda, A., Shiba, T. and Sakaki, Y.
Homo sapiens genomic DNA, chromosome 21q
Published Only in Database (1998)
2 (bases 1 to 711)
Hattori, M., Ishii, K., Toyoda, A., Shiba, T. and Sakaki, Y.
Direct Submission
Submitted (11-SEP-1998) Masahira Hattori, RIKEN Genomic Sciences
Center, RIKEN Yokohama Institute, Yokohama Research Promotion
Division, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa,
230-0045, Japan (E-mail: hattori@gsf.riken.jp, Tel: 81-45-503-9111,
Fax: 81-45-503-9113)
On Nov 16, 2004 this sequence version replaced gi:3582152.
Location/Qualifiers

1. .711
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q"
/clone="B335C24X38"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 711;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1
114
AAAAAAAAAGTTCCAATTCAGT 24
AAAAAAAAAGTTCCAATTCAGT 91

RESULT 41
AG014009
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AG014009 721 bp DNA linear GSS 16-FEB-2005
Homo sapiens genomic DNA, 21q region, clone: B335C24X38, genomic
survey sequence.
AG014009
AG014009.1 GI:3582174
GSS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Hattori, M., Ishii, K., Toyoda, A., Shiba, T. and Sakaki, Y.
Homo sapiens genomic DNA, chromosome 21q
Published Only in Database (1998)
2 (bases 1 to 721)
Hattori, M., Ishii, K., Toyoda, A., Shiba, T. and Sakaki, Y.
Direct Submission
Submitted (11-SEP-1998) Masahira Hattori, RIKEN Genomic Sciences
Center, RIKEN Yokohama Institute, Yokohama Research Promotion

Division; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa,
230-0045, Japan (E-mail: hatori@igsc.riken.jp, Tel: 81-45-503-9111,
Fax: 81-45-503-9113)

FEATURES

source

1..721
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q"
/clone="B335C24X38"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 721;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCAAATTCAGAT 24
|||||
Db 149 AAAAAAAAAAGTCCAAATTCAGAT 172

RESULT 42
DN959728/c 801 bp mRNA linear EST 09-MAY-2005
LOCUS DN959728
DEFINITION USDA-PP/ARO.15558 Star Ruby grapefruit hot water-treated flavedo
ACCESSION DN959728
VERSION DN959728.1 GI:63106462
KEYWORDS EST.
SOURCE Citrus x paradisi
ORGANISM Citrus x paradisi
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 801)
McCollum, T.G., Maul, P. and Porat, R.
Expressed Sequence Tags (ESTs) from flavedo of Star Ruby grapefruit
after hot water treatment
Unpublished (2005)
Contact: McCollum, T.G.
US Horticultural Research Laboratory
USDA, ARS
2001 S. Rock Road, Ft. Pierce, FL 34945, USA
Tel: 561-462-5836
Fax: 561-462-5986
Email: gmcollum@ushrl.ars.usda.gov
Seq primer: T3 primer

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES

source

1..801
/organism="Citrus x paradisi"
/mol_type="mRNA"
/db_xref="taxon:37656"
/clone="HW-28 C07"
/issue_type="Flavado"
/dev_stage="mature fruit"
/lab_host="SOLR"
/clone_11b="Star Ruby grapefruit hot water-treated
flavado"
/note="Organ: fruit; Vector: pBluescript II SK+; Site_1:
ECORI; Site_2: XhoI; Standard library construction
protocols from Stratagene cDNA synthesis kit (Cat No.
200401-5) and Uni-ZAP XR vector kit (Cat No. 237211)"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 801;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCAAATTCAGAT 24
|||||
Db 721 AAAAAAAAAAGTCCAAATTCAGAT 698

RESULT 43

CL101305

832 bp DNA linear GSS 05-JAN-2004

LOCUS

CL101305

ISB1-38E19_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-38E19,
genomic survey sequence.

ACCESSION

CL101305

CL101305.1 GI:40594940
GSS.
Xenopus tropicalis (western clawed frog)

REFERENCE

1 (bases 1 to 832)
Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: rsubmissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTAGTACACTATAG
Class: BAC ends
High quality sequence start: 7
High quality sequence stop: 694.

FEATURES

source

1..832
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-38E19"
/clone_11b="ISB1"
/note="Vector: pBelobAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 832;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCAAATTCAGAT 24
|||||
Db 120 AAAAAAAAAAGTCCAAATTCAGAT 143

RESULT 44
A2191471/c 838 bp DNA linear GSS 30-AUG-2000
LOCUS A2191471
DEFINITION SP 1019 B1 B06 T7A Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus
genomic clone Plate=1019 Col=11 Row=D, genomic survey sequence.
A2191471
A2191471.1 GI:8374650
GSS.
Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoidea;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 838)
Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
Swartzell, S., Wallace, J.C., Pouscka, A.J., Livingston, B.T.,
Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H.
and Hood, L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
10920195
Contact: Cameron, R.A., Davidson, E.H., Hood, L
Division of Biology 156-29

REFERENCE

A2191471
TITLE
JOURNAL
PUBMED
COMMENT

California Institute of Technology
Pasadena California 91125, USA

Tel: (626) 395-8421
Fax: (626) 793-3047

Email: acameron@caltech.edu
Plate: 1019 row: D column: 11

Seq primer: T7

Class: BAC ends

High quality sequence stop: 838.

Location/Qualifiers

1. 838

/organism="Strongylocentrotus purpuratus"

/mol_type="genomic DNA"

/db_xref="taxon:7668"

/clone="plate=1019 Col=11 Row=D"

/clone_id="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BACs.6; BAC clones in E-Coli DH10B"

ORIGIN

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 9; Length 838;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAAAAAAAAAGTTCATTCAGATA 25

563 AAAAAAAAAAGTTCATTCANATA 539

RESULT 45

CX548007

LOCUS CX548007 896 bp mRNA linear EST 12-JAN-2005
DEFINITION gmrtdNS01.16-A M13R H02.002.s3 Water stressed 48h segment 1

ACCESSION CX548007

VERSION gmrtdNS01 Glycine max cDNA 3', mRNA sequence.

KEYWORDS CX548007.1 GI:57575032

EST.

SOURCE Glycine max (soybean)

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

1 (bases 1 to 896)

REFERENCE Valliyodan,B., Huang,S., Joshi,T., Hernandez,A., Spollen,W.G.,

Bohnert,H.J., Duke,M.V., Liu,X., Scheffler,B.E., Sharp,R.E., Xu,D.,

Springer,G.K., Stacey,G. and Nguyen,H.T.

EST Analysis of Soybean Root Tip Under Drought Stress: MISC Grant

002245 - Development and Deployment of Biotechnology Tools for

Soybean Improvement

Unpublished (2005)

Contact: Henry T. Nguyen

Molecular Genetics and Biotechnology

University of Missouri-Columbia

National Center for Soybean Biotechnology, 1-87 Agriculture Bldg.,

MO 65211, USA

Tel: (573) 882-5494

Fax: (573) 882-1469

Email: nguyenhenry@missouri.edu

POLYA=yes

Location/Qualifiers

1. 896

/organism="Glycine max"

/mol_type="mRNA"

/culivar="Willam82"

/db_xref="taxon:3847"

/dev_stage="Droughted Roots"

/clone_id="gmrtdNS01"

/note="Vector: pBS II SK(+); Funding: The production of

EST's submitted in this project was funded by MISC Grant

002245 - 'Development and Deployment of Biotechnology tools

for soybean improvement' to H.T. Nguyen as Principal

Investigator. Sample collection and library construction:

Dark grown soybean seedlings with primary roots were

transplanted to high (-0.05MPa) or low water potential

(-1.6MPa) vermiculite, and harvested at 5h and 48h after

transplanting. Each root tip was divided into two segments

(distances are from the junction of root apex and root

cap): segment 1, 0-3mm; and segment 2, 3-11mm. Total RNA

was extracted by the 'hot phenol' method (Plant Molecular

Biology manual. D5: 1-13, 2nd ed., 1997). This method

worked in eliminating carbohydrate material present in the

root tips. The integrity of the RNA was verified by

denaturing agarose gels and spectrophotometry (ratio

A260/280). Poly(A)+mRNA was isolated twice from total RNA

using the Oligotex Direct RNA kit (Qiagen). Poly(A)+ mRNA

was converted to double-stranded cDNA and tagged by using

modified Oligo(dT) primers. The double stranded cDNAs were

size-selected (>500 bp). Size selected cDNAs were

adapted with EcoRI adaptors at both ends, and then

digested with NotI. The cDNA was directionally cloned into

EcoRI-NotI digested pBS II SK(+) phagemid vector

(Stratagene) and electroporated into E.coli DH10B. The

gmrtdNS01 library was created by subtracting the

well-watered soybean root library from the water-stressed

(5h and 48h) soybean root libraries. The double-stranded

plasmid DNA from the primary library was converted to

single-stranded circles and used as a template for PCR

amplification. The purified PCR products, representing the

entire cDNA population cloned were used as a driver for

normalization. Hybridization between the single-stranded

library and the PCR products was carried out for 44 hours

at 300C. Non-hybridized single-stranded DNA circles were

separated from hybridized DNA rendered partially

double-stranded and electroporated into DH10B E. coli.

Clone requests: Requests for clones should be made to Dr.

Henry Nguyen, National Center for Soybean Biotechnology,

1-87 Agriculture Bldg., University of Missouri-Columbia,

MO, 65211. Email: nguyenhenry@missouri.edu

TAG TISSUE=Water stressed 48h segment 1

TAG_SEQ=ACCGA"

ORIGIN

Query Match

Best Local Similarity 83.2%; Score 20.8; DB 8; Length 896;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 AAAAAAAAAAGTTCATTCAGATA 25

180 AAAAAAAAAAGTTCATTCAGATA 203

RESULT 46

CX548250

LOCUS CX548250 926 bp mRNA linear EST 12-JAN-2005
DEFINITION gmrtdNS01.19-A M13R E08.056.s3 Water stressed 48h segment 1

ACCESSION CX548250

VERSION gmrtdNS01 Glycine max cDNA 3', mRNA sequence.

KEYWORDS CX548250.1 GI:57575275

EST.

SOURCE Glycine max (soybean)

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

1 (bases 1 to 926)

REFERENCE Valliyodan,B., Huang,S., Joshi,T., Hernandez,A., Spollen,W.G.,

Bohnert,H.J., Duke,M.V., Liu,X., Scheffler,B.E., Sharp,R.E., Xu,D.,

Springer,G.K., Stacey,G. and Nguyen,H.T.

EST Analysis of Soybean Root Tip Under Drought Stress: MISC Grant

002245 - Development and Deployment of Biotechnology Tools for

Soybean Improvement

Unpublished (2005)

Contact: Henry T. Nguyen

Molecular Genetics and Biotechnology
University of Missouri-Columbia
National Center for Soybean Biotechnology, 1-87 Agriculture Bldg.,
MO 65211, USA
Tel: (573)882-5494
Fax: (573)882-1469
Email: nguyenhenny@missouri.edu
POLYA=yes

FEATURES

Location/Qualifiers
1..926
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Williams82"
/db_xref="taxon:3847"
/dev_stage="Droughted Roots"
/clone_1lb="gmtrDMS01"
/note="Vector: pBS II SK(+); Funding: The production of ESTs submitted in this project was funded by MSNC Grant 002245- 'Development and deployment of biotechnology tools for soybean improvement' to H.T. Nguyen as Principal Investigator. Sample collection and library construction: Dark grown soybean seedlings with primary roots were transplanted to high (-0.05MPa) or low water potential (-1.6MPa) vermiculite, and harvested at 5h and 48h after transplanting. Each root tip was divided into two segments (distances are from the junction of root apex and root cap): segment 1, 0-3mm; and segment 2, 3-11mm. Total RNA was extracted by the 'hot phenol' method (Plant Molecular Biology manual, D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct RNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. The double stranded cDNAs were size-selected (>500 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoRI-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The gmtrDMS01 library was created by subdirecting the well-watered soybean root library from the water-stressed (5h and 48h) soybean root libraries. The double stranded plasmid DNA from the primary library was converted to single-stranded circles and used as a template for PCR amplification. The purified PCR products, representing the entire cDNA population cloned were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30°C. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B E. coli. Clone requests: Requests for clones should be made to Dr. Henry Nguyen, National Center for Soybean Biotechnology, 1-87 Agriculture Bldg., University of Missouri-Columbia, MO, 65211. Email: nguyenhenny@missouri.edu
TAG TISSUE=water stressed 48h segment 1
TAG_SEQ=ACCGA"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 926;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGTTCATTTCAGATA 25
|||||
Db 180 AAAAAAAAAAGTTCATTTCAGATA 203
|||||

RESULT 47
CD509196/c 1096 bp mRNA linear EST 12-JUN-2003
LOCUS CD509196

DEFINITION CDA95-A05.5.xid-t SHGC-CDA Gasterosteus aculeatus cDNA clone
ACCESSION CDA95-A05.5, mRNA sequence.
VERSION CD509196
KEYWORDS GI:31439765
SOURCE
ORGANISM
REFERENCE
AUTHORS Kingsley,D.M., Petchel,C., Balabhadra,S., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
TITLE Expressed sequence tags from Gasterosteus aculeatus
JOURNAL Unpublished (2003)
COMMENT Contact: Kingsley, DM
HMT and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@crgm.stanford.edu
Plate: 95
High quality sequence stop: 807.

FEATURES

Location/Qualifiers
1..1096
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA95-A05"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_1lb="SHGC-CDA"
/note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with a 50 bp linker primer containing an oligo dT sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual pBK-CMV phagemid clones for EST sequencing."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 1096;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTTCAGAT 24
|||||
Db 522 AAAAAAAAAAGTTCATTTCAGAT 499
|||||

RESULT 48
CD499172 1108 bp mRNA linear EST 12-JUN-2003
LOCUS CD499172
DEFINITION CDA37-G12.3.yid-s SHGC-CDA Gasterosteus aculeatus cDNA clone
CD499172 3, mRNA sequence.
ACCESSION CD499172.1 GI:31426203
VERSION CD499172.1
KEYWORDS
SOURCE
ORGANISM
Gasterosteus aculeatus (three spined stickleback)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

REFERENCE
AUTHORS

Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1108)
Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)

TITLE
JOURNAL

HHMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cmgm.stanford.edu
Plate: 37

FEATURES

Source

High quality sequence start: 12
High quality sequence stop: 814.
Location/Qualifiers
1..1108

/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA37-G12"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/note="Vector: lambda ZAP Express/PBK-CMV; Site 1: EcoRI
(5' adaptor); Site 2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-CDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used, containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of PBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual PBK-CMV phagemid clones for
EST sequencing."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 1108;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTTCAGAT 24
|||||
Db 213 AAAAAAAAAAGTTCATTTCAGAT 236

RESULT 49
DN735647/c 1161 bp mRNA linear EST 31-MAR-2005
LOCUS
DEFINITION
CNB82-D08 y1d-s SHGC-CNB Gasterosteus aculeatus cDNA clone
CNB82-D08 5', mRNA sequence.
DN735647
ACCESSION
VERSION
DN735647.1 GI:62111883

KEYWORDS
SOURCE
ORGANISM
Gasterosteus aculeatus (three spined stickleback)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.

REFERENCE

AUTHORS

1 (bases 1 to 1161)
Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
TITLE
JOURNAL
COMMENT
Contact: Grimwood, Jane

Stanford Human Genome Center
Stanford University School of Medicine
975 S California Ave, Palo Alto, CA 94304, USA
Tel: 650 320 5917
Fax: 650 320 5801
Email: jane@hgc.stanford.edu
Plate: 82

FEATURES

Source

High quality sequence stop: 716.
Location/Qualifiers
1..1161

/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Conner Creek sticklebacks, WA USA"
/db_xref="taxon:69293"
/clone="CNB82-D08"
/sex="mixed male and female"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="SHGC-CNB"
/note="Vector: Express 1; Total and poly A+ RNA was
isolated from the indicated stickleback tissue, and a cDNA
library was constructed in the Express 1 plasmid vector by
Open Biosystems. First strand cDNA synthesis was primed
with an 54 bp linker primer containing an oligodT sequence
preceded by a synthetic NotI site (first strand primer:
5'-GACGAGTTCGATCGGAGCGGCCGCTT)25-3'). Following
second strand synthesis, cDNAs were made blunt at the end
corresponding to the original 5 prime end of mRNA, and
cloned directionally into the NotI and EcoRV sites of
Express 1. Note that the EcoRV site is typically destroyed
in the blunt end cloning, leaving a junction of the form
'xxxxATC' (where is ATC is the second half of the EcoRV
site, and xxx is derived from the cDNA sequence). A map of
the Express 1 vector is available at:
http://www.openbiosystems.com/cdna_library_construction_fa9.php#8
The primary library was transformed and amplified
in DH10B (T1 phage resistant) bacteria. Clones available
from Open Biosystems:
<http://www.openbiosystems.com/stickleback>"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 1161;
Best Local Similarity 91.7%; Pred. No. 2.6e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGTTCATTTCAGAT 24
|||||
Db 418 AAAAAAAAAAGTTCATTTCAGAT 395

RESULT 50
DN713088 1202 bp mRNA linear EST 30-MAR-2005
LOCUS
DEFINITION
CNB06-A08 x1d-c SHGC-CNB Gasterosteus aculeatus cDNA clone
CNB06-A08 3', mRNA sequence.
DN713088
ACCESSION
VERSION
DN713088.1 GI:62078127

KEYWORDS
SOURCE
ORGANISM
Gasterosteus aculeatus (three spined stickleback)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.

REFERENCE

AUTHORS

1 (bases 1 to 1202)
Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
TITLE
JOURNAL
COMMENT
Contact: Grimwood, Jane
Stanford Human Genome Center
Stanford University School of Medicine

Thu Dec 15 09:19:18 2005

us-10-681-773-8.rst

Page 24

975 S California Ave, Palo Alto, CA 94304, USA
Tel: 650 320 5917
Fax: 650 320 5801
Email: janeeshgc.stanford.edu
Place: 06
High quality sequence start: 18
High quality sequence stop: 804.

FEATURES

Source

```

/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Conner Creek sticklebacks, WA USA"
/db_xref="taxon:69293"
/clone="CNB06-A08"
/sex="mixed male and female"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="SHGC-CNB"
/notes="Vector: Express 1; Total and poly A+ RNA was isolated from the indicated stickleback tissue, and a cDNA library was constructed in the Express 1 plasmid vector by Open Biosystems. First strand cDNA synthesis was primed with an 54 bp linker primer containing an oligot sequence preceded by a synthetic NotI site (first strand primer: 5'-GACAGATTTCAGATCGACGACGCCGCC(T)25-3'). Following second strand synthesis, cDNAs were made blunt at the end corresponding to the original 5' prime end of mRNA, and cloned directionally into the NotI and EcoRV sites of Express 1. Note that the EcoRV site is typically destroyed in the blunt end cloning, leaving a junction of the form 5'...xxxxATC(where is ATC is the second half of the EcoRV site, and xxx is derived from the cDNA sequence). A map of the Express 1 vector is available at:
http://www.openbiosystems.com/cdna_library_constructionfaq.php# The primary library was transformed and amplified in DH10B (T1 phage resistant) bacteria. Clones available from Open Biosystems:
http://www.openbiosystems.com/stickleback"

```

ORIGIN

Query Match	83.2%	Score 20.8;	DB 8;	Length 1202;
Best Local Similarity	91.7%	Pred. No. 2.6e+03;		
Matches 22;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps 0;

[illegible]

Search completed: December 14, 2005, 07:35:28
Job time : 1759.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:43:33 ; Search time 180.2 Seconds
(without alignments)
68.002 Million cell updates/sec

Title: US-10-681-773-8

Perfect score: 25
Sequence: 1 aaaaaaaaaagcttcacatcagata 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4161359 seqs, 245077644 residues

Total number of hits satisfying chosen parameters: 8322718

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

Published Applications NA New:
1: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US12_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US13_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US14_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US15_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US16_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	20.2	80.8	398287	6 US-10-995-561-13396	Sequence 13396, A
2	19.2	76.8	1005	6 US-10-750-185-51843	Sequence 51843, A
3	19.2	76.8	1173	6 US-10-750-185-51549	Sequence 45149, A
4	19.2	76.8	150314	7 US-11-112-908-24	Sequence 24, Appl
5	18.8	75.2	201	6 US-10-995-561-42775	Sequence 42775, A
6	18.8	75.2	201	6 US-10-995-561-44105	Sequence 44105, A
7	18.8	75.2	1123000	6 US-10-995-561-13286	Sequence 13286, A
8	18.6	74.4	201	6 US-10-995-561-71256	Sequence 71256, A
9	18.6	74.4	909	6 US-10-750-185-51557	Sequence 41557, A
10	18.6	74.4	1302	6 US-10-750-185-51597	Sequence 35907, A
11	18.6	74.4	2118	6 US-10-750-185-51597	Sequence 35907, A
12	18.6	74.4	2174	6 US-10-750-185-51597	Sequence 41597, A
13	18.6	74.4	3571	6 US-10-750-185-51302	Sequence 31302, A
14	18.6	74.4	86585	7 US-11-112-908-198	Sequence 198, App
15	18.6	74.4	148935	6 US-10-995-561-13308	Sequence 13308, A
16	18.6	74.4	155989	7 US-11-121-086-57	Sequence 57, Appl
17	18.6	74.4	180654	7 US-11-121-086-58	Sequence 58, Appl
18	18.6	74.4	212716	7 US-11-121-086-95	Sequence 95, Appl
19	18.6	74.4	403278	6 US-10-995-561-13421	Sequence 13421, A
20	18.6	73.6	63984	7 US-11-121-086-26	Sequence 26, Appl
21	18.2	72.8	201	6 US-10-995-561-43911	Sequence 43911, A
22	18.2	72.8	201	6 US-10-995-561-43912	Sequence 43912, A
23	18.2	72.8	201	6 US-10-995-561-43913	Sequence 43913, A

24	18.2	72.8	201	6 US-10-995-561-43914	Sequence 43914, A
25	18.2	72.8	201	6 US-10-995-561-66191	Sequence 66191, A
26	18.2	72.8	201	6 US-10-995-561-66192	Sequence 66192, A
27	18.2	72.8	201	6 US-10-995-561-66193	Sequence 66193, A
28	18.2	72.8	201	6 US-10-995-561-66194	Sequence 66194, A
29	18.2	72.8	886	6 US-10-750-185-37910	Sequence 37910, A
30	18.2	72.8	1056	6 US-10-750-185-37999	Sequence 37999, A
31	18.2	72.8	1799	6 US-10-750-185-45222	Sequence 45222, A
32	18.2	72.8	2125	6 US-10-750-185-54826	Sequence 54826, A
33	18.2	72.8	13532	6 US-10-995-561-13301	Sequence 13301, A
34	18.2	72.8	13597	6 US-10-995-561-13302	Sequence 13302, A
35	17.8	71.2	201	6 US-10-995-561-32107	Sequence 32107, A
36	17.8	71.2	991	6 US-10-995-561-13237	Sequence 13237, A
37	17.8	71.2	87672	6 US-10-995-561-64268	Sequence 64268, A
38	17.8	71.2	103650	6 US-10-995-561-13253	Sequence 13253, A
39	17.8	71.2	117431	6 US-10-995-561-13253	Sequence 13253, A
40	17.8	71.2	235033	7 US-11-157-389-1	Sequence 1, Appl
41	17.8	71.2	237326	7 US-11-157-389-2	Sequence 2, Appl
42	17.8	71.2	285300	6 US-10-995-561-13448	Sequence 13448, A
43	17.8	71.2	340000	6 US-10-995-561-13448	Sequence 13448, A
44	17.6	70.4	201	6 US-10-995-561-18280	Sequence 18280, A
45	17.6	70.4	201	6 US-10-995-561-27552	Sequence 27552, A
46	17.6	70.4	201	6 US-10-995-561-50068	Sequence 50068, A
47	17.6	70.4	201	6 US-10-995-561-64267	Sequence 64267, A
48	17.6	70.4	201	6 US-10-995-561-64268	Sequence 64268, A
49	17.6	70.4	201	6 US-10-995-561-64268	Sequence 64268, A
50	17.6	70.4	552	6 US-10-750-185-1781	Sequence 1781, Ap
51	17.6	70.4	714	6 US-10-750-185-45969	Sequence 45969, A
52	17.6	70.4	809	6 US-10-750-185-46095	Sequence 46095, A
53	17.6	70.4	838	6 US-10-750-185-54804	Sequence 54804, A
54	17.6	70.4	1036	6 US-10-750-185-53216	Sequence 53216, A
55	17.6	70.4	1037	6 US-10-750-185-62080	Sequence 62080, A
56	17.6	70.4	1132	6 US-10-750-185-53964	Sequence 53964, A
57	17.6	70.4	1131	6 US-10-750-185-53367	Sequence 53367, A
58	17.6	70.4	2153	6 US-10-750-185-59199	Sequence 59199, A
59	17.6	70.4	2934	6 US-10-750-185-41151	Sequence 41151, A
60	17.6	70.4	3791	7 US-11-093-746A-1	Sequence 1, Appl
61	17.6	70.4	18930	6 US-10-995-561-13213	Sequence 13213, A
62	17.6	70.4	43436	6 US-10-995-561-13240	Sequence 13240, A
63	17.6	70.4	56448	6 US-10-995-561-13369	Sequence 13369, A
64	17.6	70.4	118996	7 US-11-121-086-84	Sequence 84, Appl
65	17.6	70.4	138821	7 US-11-121-086-80	Sequence 80, Appl
66	17.6	70.4	139054	7 US-11-121-086-96	Sequence 96, Appl
67	17.6	70.4	201950	6 US-10-995-561-13403	Sequence 13403, A
68	17.6	70.4	394468	6 US-10-995-561-13473	Sequence 13473, A
69	17.6	70.4	1082144	7 US-11-117-187-211	Sequence 211, App
70	17.4	69.6	201	6 US-10-995-561-20830	Sequence 20830, A
71	17.4	69.6	1510	6 US-10-750-185-53214	Sequence 53214, A
72	17.4	69.6	44718	6 US-10-995-561-13217	Sequence 13217, A
73	17.4	69.6	150450	7 US-11-112-908-15	Sequence 15, Appl
74	17.4	69.6	169047	7 US-11-121-086-15	Sequence 15, Appl
75	17.4	69.6	191343	7 US-11-121-086-15	Sequence 15, Appl
76	17.4	69.6	321019	6 US-10-995-561-13204	Sequence 13204, A
77	17.2	68.8	201	6 US-10-995-561-58573	Sequence 58573, A
78	17.2	68.8	600	6 US-10-750-185-208	Sequence 208, App
79	17.2	68.8	1461	6 US-10-750-185-49881	Sequence 49881, A
80	17.2	68.8	1501	6 US-10-750-185-4797	Sequence 4797, A
81	17.2	68.8	1799	6 US-10-750-185-27608	Sequence 27608, A
82	17.2	68.8	1819	6 US-10-955-054-35	Sequence 35, Appl
83	17.2	68.8	2421	6 US-10-750-185-27683	Sequence 27683, A
84	17.2	68.8	12537	6 US-10-995-561-13224	Sequence 13224, A
85	17.2	68.8	59110	6 US-10-995-561-1324	Sequence 13235, A
86	17.2	68.8	105550	6 US-10-995-561-13335	Sequence 68, Appl
87	17.2	68.8	146656	7 US-11-121-086-68	Sequence 68, Appl
88	17.2	68.8	154548	7 US-11-121-086-33	Sequence 33, Appl
89	17.2	68.8	177623	7 US-11-112-908-41	Sequence 41, Appl
90	17.2	68.8	191353	7 US-11-112-908-51	Sequence 51, Appl
91	17	68.0	181	6 US-10-623-155-217	Sequence 217, App
92	17	68.0	201	6 US-10-995-561-26667	Sequence 26667, A
93	17	68.0	201	6 US-10-995-561-39704	Sequence 39704, A
94	17	68.0	201	6 US-10-995-561-40350	Sequence 40350, A
95	17	68.0	201	6 US-10-995-561-46511	Sequence 46511, A
96	17	68.0	201	6 US-10-995-561-50930	Sequence 50930, A

```
C 97 17 68.0 201 6 US-10-995-561-50966 Sequence 50966, A
C 98 17 68.0 201 6 US-10-995-561-50967 Sequence 50967, A
C 99 17 68.0 201 6 US-10-995-561-50977 Sequence 50977, A
C 100 17 68.0 201 6 US-10-995-561-50980 Sequence 50980, A
C 101 17 68.0 201 6 US-10-995-561-51067 Sequence 51067, A
C 102 17 68.0 201 6 US-10-995-561-51096 Sequence 51096, A
C 103 17 68.0 201 6 US-10-995-561-51914 Sequence 51914, A
C 104 17 68.0 201 6 US-10-995-561-58868 Sequence 58868, A
C 105 17 68.0 201 6 US-10-995-561-58870 Sequence 58870, A
C 106 17 68.0 201 6 US-10-995-561-58872 Sequence 58872, A
C 107 17 68.0 201 6 US-10-995-561-57542 Sequence 67542, A
C 108 17 68.0 201 6 US-10-995-561-67885 Sequence 67885, A
C 109 17 68.0 201 6 US-10-995-561-67897 Sequence 67897, A
C 110 17 68.0 201 6 US-10-995-561-74322 Sequence 74322, A
C 111 17 68.0 600 6 US-10-750-185-759 Sequence 759, App
C 112 17 68.0 600 6 US-10-750-185-1503 Sequence 1503, App
C 113 17 68.0 600 6 US-10-750-185-2376 Sequence 2376, App
C 114 17 68.0 600 6 US-10-750-185-2643 Sequence 2643, App
C 115 17 68.0 652 6 US-10-750-185-32071 Sequence 32071, A
C 116 17 68.0 743 6 US-10-750-185-58040 Sequence 58040, A
C 117 17 68.0 768 6 US-10-750-185-58325 Sequence 58325, A
C 118 17 68.0 827 6 US-10-750-185-58701 Sequence 58701, A
C 119 17 68.0 842 6 US-10-750-185-37787 Sequence 37787, A
C 120 17 68.0 847 6 US-10-750-185-52573 Sequence 52573, A
C 121 17 68.0 1064 6 US-10-750-185-26059 Sequence 26059, A
C 122 17 68.0 1191 6 US-10-750-185-61744 Sequence 61744, A
C 123 17 68.0 1227 6 US-10-750-185-63598 Sequence 63598, A
C 124 17 68.0 1252 6 US-10-750-185-62619 Sequence 62619, A
C 125 17 68.0 1254 6 US-10-750-185-34720 Sequence 34720, A
C 126 17 68.0 1256 6 US-10-750-185-53641 Sequence 53641, A
C 127 17 68.0 1267 6 US-10-750-185-36648 Sequence 36648, A
C 128 17 68.0 1274 6 US-10-750-185-36648 Sequence 36648, A
C 129 17 68.0 1325 6 US-10-750-185-26491 Sequence 26491, A
C 130 17 68.0 1325 6 US-10-750-185-28744 Sequence 28744, A
C 131 17 68.0 1355 6 US-10-750-185-43944 Sequence 43944, A
C 132 17 68.0 1364 6 US-10-750-185-53256 Sequence 53256, A
C 133 17 68.0 1516 6 US-10-750-185-27515 Sequence 27515, A
C 134 17 68.0 1533 6 US-10-750-185-53789 Sequence 53789, A
C 135 17 68.0 1543 6 US-10-750-185-39124 Sequence 39124, A
C 136 17 68.0 1618 6 US-10-750-185-56437 Sequence 56437, A
C 137 17 68.0 1655 6 US-10-750-185-53109 Sequence 53109, A
C 138 17 68.0 1691 6 US-10-750-185-26289 Sequence 26289, A
C 139 17 68.0 1736 6 US-10-750-185-42557 Sequence 42557, A
C 140 17 68.0 1764 6 US-10-750-185-56585 Sequence 56585, A
C 141 17 68.0 1858 6 US-10-750-185-24514 Sequence 24514, A
C 142 17 68.0 1873 6 US-10-750-185-36272 Sequence 36272, A
C 143 17 68.0 1886 6 US-10-750-185-44675 Sequence 44675, A
C 144 17 68.0 1953 6 US-10-750-185-44475 Sequence 44475, A
C 145 17 68.0 2046 6 US-10-750-185-63693 Sequence 63693, A
C 146 17 68.0 2243 6 US-10-750-185-25482 Sequence 25482, A
C 147 17 68.0 2269 6 US-10-750-185-25238 Sequence 25238, A
C 148 17 68.0 2498 6 US-10-750-185-34470 Sequence 34470, A
C 149 17 68.0 2555 6 US-10-750-185-34470 Sequence 34470, A
C 150 17 68.0 2759 6 US-10-750-185-34470 Sequence 34470, A
```

ALIGNMENTS

```
RESULT 1
US-10-995-561-13396/C
; Sequence 13396, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: PASTSEQ for Windows Version 4.0
```

```
; SEQ ID NO 13396
; LENGTH: 398287
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1...398287)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13396
```

```
Query Match 80.8%; Score 20.2; DB 6; Length 398287;
Best Local Similarity 88.0%; Pred. No. 33;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 229943 AAAAAAAAAAGTTCATTCATATA 229919
```

```
RESULT 2
US-10-750-185-51843/C
; Sequence 51843, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51843
; LENGTH: 1005
; TYPE: DNA
; ORGANISM: Bovine 19866881513039
US-10-750-185-51843
```

```
Query Match 76.8%; Score 19.2; DB 6; Length 1005;
Best Local Similarity 87.5%; Pred. No. 83;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 770 AAAAAACAAAATTCCAATCAGAT 747
```

```
RESULT 3
US-10-750-185-45149/C
; Sequence 45149, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
```

SOFTWARE: Patentin version 3.1
SEQ ID NO 45149
LENGTH: 1173
TYPE: DNA
ORGANISM: Bovine 1986881021600
US-10-750-185-45149

Query Match 76.8%; Score 19.2; DB 6; Length 1173;
Best Local Similarity 87.5%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAGTTCATTCAGAT 25
Db 964 AAAAAAGCTTCATTCAGT 941

RESULT 4
US-11-112-908-24/c
Sequence 24, Application US/11112908
Publication No. US20050260659A1
GENERAL INFORMATION:
APPLICANT: Harris, Cole
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: Patentin version 3.3
SEQ ID NO 24
LENGTH: 150314
TYPE: DNA
ORGANISM: Homo sapiens
US-11-112-908-24

Query Match 76.8%; Score 19.2; DB 7; Length 150314;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAGAT 24
Db 66643 AAAAAAGCTTTAACTCAGAT 66620

RESULT 5
US-10-995-561-42775
Sequence 42775, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: C1001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 42775
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-42775

Query Match 75.2%; Score 18.8; DB 6; Length 201;

Best Local Similarity 90.9%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAGTTCATTCAG 22
Db 110 AAAAAAGCTTTAAATCAG 131

RESULT 6
US-10-995-561-44105
Sequence 44105, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: C1001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 44105
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-44105

Query Match 75.2%; Score 18.8; DB 6; Length 201;
Best Local Similarity 90.9%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAG 22
Db 109 AAAAAAGCTTTAAATCAG 130

RESULT 7
US-10-995-561-13286
Sequence 13286, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: C1001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13286
LENGTH: 1125000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(1125000)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13286

Query Match 75.2%; Score 18.8; DB 6; Length 1125000;
Best Local Similarity 90.9%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAGTTCATTCAG 22
Db 541068 AAAAAAGCTTTAAATCAG 541089

RESULT 8
US-10-995-561-71256
Sequence 71256, Application US/10995561

```
/ Publication No. US20050272054A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
/ FILE REFERENCE: CL001559
/ CURRENT APPLICATION NUMBER: US/10/995,561
/ CURRENT FILING DATE: 2004-11-24
/ NUMBER OF SEQ ID NOS: 85702
/ SOFTWARE: PasteSeq for Windows Version 4.0
/ SEQ ID NO 71256
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-995-561-71256

Query Match          74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 92 AAAAAAAAAAGTACCAATAAAGATA 116

RESULT 9
US-10-750-185-41557/c
/ Sequence 41557, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 41557
/ LENGTH: 909
/ TYPE: DNA
/ ORGANISM: Bovine
US-10-750-185-41557

Query Match          74.4%; Score 18.6; DB 6; Length 909;
Best Local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 180 AAAAAAAAAAGTCCATTTCAGATA 156

RESULT 10
US-10-750-185-35907/c
/ Sequence 35907, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
```

```
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 35907
/ LENGTH: 1302
/ TYPE: DNA
/ ORGANISM: Bovine
US-10-750-185-35907

Query Match          74.4%; Score 18.6; DB 6; Length 1302;
Best Local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 1065 AAACAAAAAAGTCCAATTAAAGACA 1041

RESULT 11
US-10-750-185-49665
/ Sequence 49665, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 49665
/ LENGTH: 2118
/ TYPE: DNA
/ ORGANISM: Bovine
US-10-750-185-49665

Query Match          74.4%; Score 18.6; DB 6; Length 2118;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAAGTCCATTTCAGATA 25
Db 505 AAAAAAAAAAGTACGTTTCAGTTA 529

RESULT 12
US-10-750-185-31302/c
/ Sequence 31302, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM1100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
```

```

; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31302
; LENGTH: 2174
; TYPE: DNA
; ORGANISM: Bovine 19866881109047
US-10-750-185-31302

Query Match          74.4% Score 18.6; DB 6; Length 2174;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 1617 AAAAAAAAAATTCATTCAGATA 1593

RESULT 13
US-10-750-185-33894/c
; Sequence 33894, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33894
; LENGTH: 3571
; TYPE: DNA
; ORGANISM: Bovine 19866881061705
US-10-750-185-33894

Query Match          74.4% Score 18.6; DB 6; Length 3571;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 2101 AAAAAAAAAAGTTAGATTGAGATA 2077

RESULT 14
US-11-117-187-198
; Sequence 198, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPEHNAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 198
```

```

; LENGTH: 86585
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-11-117-187-198

Query Match          74.4% Score 18.6; DB 7; Length 86585;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 32095 AAAAAAAAAAGTTAAATTCAGAGA 32119

RESULT 15
US-10-995-561-13308/c
; Sequence 13308, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CU001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13308
; LENGTH: 148935
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13308

Query Match          74.4% Score 18.6; DB 6; Length 148935;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 86942 AAAAAAAAAAGTTCAGTTCCAGAGA 86918

RESULT 16
US-11-121-086-57/c
; Sequence 57, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 57
; LENGTH: 155989
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-57

Query Match          74.4% Score 18.6; DB 7; Length 155989;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAGTTCATTCAGATA 25
Db 5717 AAAAAAAAAATTCATTCAGATA 5693
```

```
RESULT 17
; US-11-121-086-58/c
; Sequence 58, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 58
; LENGTH: 180654
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-58

Query Match          74.4%; Score 18.6; DB 7; Length 180654;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCCAATTCAGATA 25
Db 165867 AAAAAATATAATTCCAATTCGATA 165843

RESULT 18
; US-11-121-086-95/c
; Sequence 95, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 95
; LENGTH: 212716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-95

Query Match          74.4%; Score 18.6; DB 7; Length 212716;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCCAATTCAGATA 25
Db 60083 AAACAAAAGGTACAAATTCAGATA 60059

RESULT 19
; US-10-995-561-13421
; Sequence 13421, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13421
; LENGTH: 403278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(403278)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13421

Query Match          74.4%; Score 18.6; DB 6; Length 403278;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCCAATTCAGATA 25
Db 12631 AAAAAAAAAAGTTCCAATTAAGATA 12655

RESULT 20
; US-11-121-086-26/c
; Sequence 26, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 63984
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-26

Query Match          73.6%; Score 18.4; DB 7; Length 63984;
Best Local Similarity 95.0%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCCAATTC 20
Db 30230 AAAAAAAAAAGTTCATTC 30211

RESULT 21
; US-10-995-561-43911
; Sequence 43911, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43911
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-43911

Query Match          72.8%; Score 18.2; DB 6; Length 201;
Best Local Similarity 87.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Query Match	72.8%	Score	18.2	DB	6	Length	201
Best Local Similarity	87.0%	Pred. No.	1.7e+02				
Matches	20	Conservative	0	Mismatches	3	Indels	0
				Gaps	0		0

QY 1 AAAAAAAAAAGTCCATTGACA 23
 Db 93 AAAAAAAAAAGTCCATTGACA 115

RESULT 27
 US-10-995-561-66193
 ; Sequence 66193, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 66193
 ; LENGTH: 201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-995-561-66193

Query Match 72.8%; Score 18.2; DB 6; Length 201;
 Best Local Similarity 87.0%; Pred. No. 1.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACA 23
 Db 94 AAAAAAAAAAGTCCATTGACA 116

RESULT 28
 US-10-995-561-66194
 ; Sequence 66194, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 66194
 ; LENGTH: 201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-995-561-66194

Query Match 72.8%; Score 18.2; DB 6; Length 201;
 Best Local Similarity 87.0%; Pred. No. 1.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACA 23
 Db 95 AAAAAAAAAAGTCCATTGACA 117

RESULT 29
 US-10-750-185-37910
 ; Sequence 37910, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom

APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM11100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 37910
 ; LENGTH: 886
 ; TYPE: DNA
 ; ORGANISM: Bovine
 US-10-750-185-37910

Query Match 72.8%; Score 18.2; DB 6; Length 886;
 Best Local Similarity 87.0%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCATTGACA 23
 Db 53 AAAAAAAAAAGTCCATTGACA 75

RESULT 30
 US-10-750-185-37999
 ; Sequence 37999, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM11100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 37999
 ; LENGTH: 1056
 ; TYPE: DNA
 ; ORGANISM: Bovine
 US-10-750-185-37999

Query Match 72.8%; Score 18.2; DB 6; Length 1056;
 Best Local Similarity 87.0%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAGTCCATTGACATA 25
 Db 1010 AAAAAAAAAAGTCCATTGACATA 1032

RESULT 31
 US-10-750-185-45222
 ; Sequence 45222, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS


```
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 45222
LENGTH: 1799
TYPE: DNA
ORGANISM: Bovine 1986680938368
US-10-750-185-45222

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 6; Length 1799;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy
1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 1585 AAAAAAAAAATCTTCAATTCAGA 1607

RESULT 32
US-10-750-185-54826/c
Sequence 54826, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 54826
LENGTH: 2125
TYPE: DNA
ORGANISM: Bovine 1986680938781
US-10-750-185-54826

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 6; Length 2125;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy
1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 1650 AAAAAAAAAATTACCAATTCAGA 1628

RESULT 33
US-10-995-561-13382
Sequence 13382, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CLO01559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13382
LENGTH: 12532

TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-13382

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 6; Length 12532;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy
1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 7361 AAAAAAAAAAGTCCCAATTCAGA 7383

RESULT 34
US-10-995-561-13301/c
Sequence 13301, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: CLO01559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13301
LENGTH: 35997
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(35997)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13301

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 6; Length 35997;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy
1 AAAAAAAAAAGTCCCAATTCAGA 23
Db 24032 AAAAAAAAAAGTTAAATTAAGA 24010

RESULT 35
US-10-995-561-31099/c
Sequence 31099, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CLO01559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31099
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-31099

Query Match
Best Local Similarity 71.2%; Score 17.8; DB 6; Length 201;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy
1 AAAAAAAAAAGTCCCAATTCAGA 21
Db 108 AAAAAAAAAAGTCCCAATTCAGA 88
```

```
RESULT 36
US-10-750-185-32107/c
; Sequence 32107, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32107
; LENGTH: 991
; TYPE: DNA
; ORGANISM: Bovine 19866881047150
US-10-750-185-32107

Query Match 71.2%; Score 17.8; DB 6; Length 991;
Best Local Similarity 90.5%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 21
DB 521 AAAAAAAAAAGTCCCAATTCA 501

RESULT 37
US-10-995-561-13237
; Sequence 13237, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CU001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13237
; LENGTH: 87672
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13237

Query Match 71.2%; Score 17.8; DB 6; Length 87672;
Best Local Similarity 90.5%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 21
DB 45884 AAAAAAAAAAGTCCCAATTCA 45904

RESULT 38
US-10-995-561-13253/c
; Sequence 13253, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CU001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13253
; LENGTH: 103660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(103660)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-995-561-13253

Query Match 71.2%; Score 17.8; DB 6; Length 103660;
Best Local Similarity 90.5%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 21
DB 32389 AAAAAAAAAAGTCCCAATTCA 32369

RESULT 39
US-10-995-561-13448/c
; Sequence 13448, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CU001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13448
; LENGTH: 117431
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13448

Query Match 71.2%; Score 17.8; DB 6; Length 117431;
Best Local Similarity 90.5%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTCCCAATTCA 21
DB 12515 AAAAAAAAAAGTCCCAATTCA 12495

RESULT 40
US-11-157-389-1/c
; Sequence 1, Application US/11157389
; Publication No. US20050266481A1
; GENERAL INFORMATION:
; APPLICANT: Ruddy, David A.
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: POLYMORPHISMS IN THE REGION OF THE HUMAN
; TITLE OF INVENTION: HEMOCHROMATOSIS GENE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/11/157,389
FILING DATE: 20-June-2005
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/852,495
FILING DATE: 07-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0057-999
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 235033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-11-157-389-1

Query Match      71.2%  Score 17.8; DB 7; Length 235033;
Best Local Similarity 90.5%  Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 AAAAAAAAAAGTTCATTC A 21
Db      187050 AAAAAAAAAAGTTCATTC A 187030

RESULT 41
US-11-157-389-2/c
Sequence 2, Application US/11157389
Publication No. US20050266481A1
GENERAL INFORMATION:
APPLICANT: Ruddy, David A.
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: POLYMORPHISMS IN THE REGION OF THE HUMAN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/11/157,389
FILING DATE: 20-June-2005
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/852,495
FILING DATE: 07-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0057-999
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 237326 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-11-157-389-2

Query Match      71.2%  Score 17.8; DB 7; Length 237326;
Best Local Similarity 90.5%  Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 AAAAAAAAAAGTTCATTC A 21
Db      189274 AAAAAAAAAAGTTCATTC A 189254

RESULT 42
US-10-857-780-6/c
Sequence 6, Application US/10857780
Publication No. US20050272043A1
GENERAL INFORMATION:
APPLICANT: ROTH, RICHARD B.
APPLICANT: BRAUN, ANDREAS
APPLICANT: KAMMERER, STEFAN M.
APPLICANT: NELSON, MATTHEW ROBERTS
APPLICANT: REINELAND, RIKARD HENRY
APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
FILE REFERENCE: SEQ-4069-CP
CURRENT APPLICATION NUMBER: US/10/857,780
CURRENT FILING DATE: 2004-05-28
PRIOR APPLICATION NUMBER: 10/723,681
PRIOR FILING DATE: 2003-11-25
PRIOR APPLICATION NUMBER: 60/490,234
PRIOR FILING DATE: 2003-07-24
PRIOR APPLICATION NUMBER: 60/525,239
PRIOR FILING DATE: 2003-11-25
NUMBER OF SEQ ID NOS: 4962
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 285300
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (3185)..(3185)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (7922)..(7922)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (10272)..(10272)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (28873)..(28873)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (29105)..(29106)
OTHER INFORMATION: n is a, c, g, or t

```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (35901)..(35901)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36488)..(36488)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40260)..(40260)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (52594)..(52594)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (68230)..(68230)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (70027)..(70027)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (83514)..(83514)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (89477)..(89477)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (108762)..(108762)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (109378)..(109378)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (113639)..(113639)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (115312)..(115312)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (115315)..(115315)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (127817)..(127817)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (138224)..(138224)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140476)..(140476)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147488)..(147488)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147500)..(147500)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
```

```
NAME/KEY: misc_feature
LOCATION: (147527)..(147527)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147534)..(147534)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (171395)..(171395)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (195657)..(195657)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (196415)..(196415)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (196417)..(196417)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (212236)..(212237)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (212442)..(212442)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (212445)..(212445)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (212455)..(212455)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (215163)..(215163)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (217657)..(217657)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (217725)..(217725)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (231757)..(231757)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (231762)..(231762)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (235827)..(235827)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (241414)..(241414)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (248915)..(248915)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
```

```

; LOCATION: (250079) .. (250082)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (250575) .. (250575)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature

Query Match      71.2%; Score 17.8; DB 6; Length 285300;
Best Local Similarity 90.5%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGTCCATTCA 21
Db      17534 AAAAAAAAAAGTCCATTCA 17514

RESULT 43
US-11-102-978-3
; Sequence 3, Application US/11102978
; Publication No. US20050250142A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Research Foundation
; TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
; FILE REFERENCE: 0274-5537.1US
; CURRENT APPLICATION NUMBER: US/11/102,978
; PRIOR FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: PCT/US2003/033152
; PRIOR FILING DATE: 2003-10-18
; PRIOR APPLICATION NUMBER: 60/419,576
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 340000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (56948) .. (57115)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8006) .. (81089)
; OTHER INFORMATION: Gene VDAC2P; voltage-dependent anion channel isoform 2 pseudogene
; FEATURE:
; NAME/KEY: exon
; LOCATION: (167308) .. (167438)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: exon
; LOCATION: (216732) .. (216833)
; OTHER INFORMATION: C21orf34 exon
; US-11-102-978-3

Query Match      71.2%; Score 17.8; DB 7; Length 340000;
Best Local Similarity 90.5%; Pred. No. 3.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGTCCATTCA 21
Db      291322 AAAAAAAAAAGTCCATTCA 291342

RESULT 44
US-10-995-561-18280
; Sequence 18280, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50068
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-995-561-18280

Query Match      70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGTCCATTCCAGAT 24
Db      21 AAAAAAAAAAGTCCATTCCAT 44

RESULT 45
US-10-995-561-27552
; Sequence 27552, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27552
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-995-561-27552

Query Match      70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGTCCATTCCAGAT 24
Db      92 AAAAAAAAAAGTCCATTCCAGCT 115

RESULT 46
US-10-995-561-50068
; Sequence 50068, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50068
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-995-561-50068

Query Match      70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 AAAAAAAAAAGTCCATTCCAGAT 24
Db      92 AAAAAAAAAAGTCCATTCCAGCT 115
```

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
 Db 134 AAAAAAAAAAGTTCATTCAGAT 157

RESULT 47

US-10-995-561-64267
 ; Sequence 64267, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 64267
 ; LENGTH: 201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-995-561-64267

Query Match 70.4%; Score 17.6; DB 6; Length 201;
 Best Local Similarity 83.3%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
 Db 73 AAAAAAAAAAGTTCATTCAGAT 96

RESULT 48

US-10-995-561-64268
 ; Sequence 64268, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 64268
 ; LENGTH: 201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-995-561-64268

Query Match 70.4%; Score 17.6; DB 6; Length 201;
 Best Local Similarity 83.3%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
 Db 44 AAAAAAAAAAGTTCATTCAGAT 67

RESULT 49

US-10-995-561-64483
 ; Sequence 64483, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF

; TITLE OF INVENTION: DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001559

; CURRENT APPLICATION NUMBER: US/10/995,561
 ; CURRENT FILING DATE: 2004-11-24
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 64483
 ; LENGTH: 201
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-995-561-64483

Query Match 70.4%; Score 17.6; DB 6; Length 201;
 Best Local Similarity 83.3%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
 Db 2 AAAAAAAAAAGTTCATTCAGAT 25

RESULT 50

US-10-750-185-1781
 ; Sequence 1781, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: DENISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1781
 ; LENGTH: 552
 ; TYPE: DNA
 ; ORGANISM: Bovine NMRI18896
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)-(1)
 ; OTHER INFORMATION: n is any nucleotide
 US-10-750-185-1781

Query Match 70.4%; Score 17.6; DB 6; Length 552;
 Best Local Similarity 83.3%; Pred. No. 3.1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGTTCATTCAGAT 24
 Db 48 AAAAAAAAAAGTTCATTCAGAT 71

Search completed: December 14, 2005, 11:40:51
 Job time : 186.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: us-10-681-773-9

Sequence: 1 aaaaaaaaaacattcatcatttaaac 25

Scoring table: IDENTITY NUC
Gapop 10'-0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23.4	93.6	719	10	BV550550 S216P6406
2	23.4	93.6	130530	8	AF131217 Homo sapi
3	23.4	93.6	340000	8	HS21C047
4	22.4	89.6	146608	14	AC114719
5	22.4	89.6	237844	14	AC098628
6	21.8	87.2	480	6	AX397279
7	21.8	87.2	30740	14	BV522543
8	21.8	87.2	38531	8	AC006129
9	21.8	87.2	39233	8	AC005626
10	21.8	87.2	110000	14	AC111566
11	21.8	87.2	110000	14	AC128588_0
12	21.8	87.2	110000	14	CR555291_0
13	21.8	87.2	117322	5	AL672217
14	21.8	87.2	134614	8	BS000225
15	21.8	87.2	141776	15	AC148528
16	21.8	87.2	142215	8	BS000165
17	21.8	87.2	171172	5	EX323586
18	21.8	87.2	171172	5	EX323586

C	19	21.8	87.2	171984	14	AC164486	AC164486 Danio rerio	
C	20	21.8	87.2	191493	14	AC118168	AC118168 Rattus norvegicus	
C	21	21.8	87.2	217987	5	CR792443	CR792443 Zebrafish	
C	22	21.8	87.2	223743	14	AC098989	AC098989 Rattus norvegicus	
C	23	21.8	87.2	231490	5	CR854945	CR854945 Zebrafish	
C	24	21.8	87.2	300486	14	AC129639	AC129639 Rattus norvegicus	
C	25	21.4	85.6	147902	5	BX548026	BX548026 Zebrafish	
C	26	21	84.0	202842	9	AL607108	AL607108 Mouse DNA	
C	27	20.8	83.2	5203	6	AX345821	AX345821 Sequence	
C	28	20.8	83.2	29482	2	CEM05B5	CEM05B5 Zebrafish	
C	29	20.8	83.2	35396	2	U00046	U00046 Caenorhabditis elegans	
C	30	20.8	83.2	37487	6	AX655389	AX655389 Sequence	
C	31	20.8	83.2	66440	14	AC016834	AC016834 Homo sapiens	
C	32	20.8	83.2	79374	8	AL162491	AL162491 Human DNA	
C	33	20.8	83.2	100161	14	AC164748	AC164748 Bos taurus	
C	34	20.8	83.2	101261	14	AP007393	AP007393 Lotus corniculatus	
C	35	20.8	83.2	110000	14	AC109967_0	AC109967 Rattus norvegicus	
C	36	20.8	83.2	112085	8	AL133544	AL133544 Human DNA	
C	37	20.8	83.2	118591	14	AC020202	AC020202 Drosophila melanogaster	
C	38	20.8	83.2	122884	8	HS1158B12	HS1158B12 Human DNA	
C	39	20.8	83.2	123647	2	AC007837	AC007837 Drosophila melanogaster	
C	40	20.8	83.2	125793	5	BX942828	BX942828 Zebrafish	
C	41	20.8	83.2	136730	8	AC108110	AC108110 Homo sapiens	
C	42	20.8	83.2	137336	15	AC007915	AC007915 Genomic sequence	
C	43	20.8	83.2	141674	8	AC010361	AC010361 Homo sapiens	
C	44	20.8	83.2	144964	14	AC162580	AC162580 Bos taurus	
C	45	20.8	83.2	146043	14	CR548641	CR548641 Danio rerio	
C	46	20.8	83.2	148196	8	BS000106	BS000106 Pan troglodytes	
C	47	20.8	83.2	156485	8	AC007422	AC007422 Homo sapiens	
C	48	20.8	83.2	156753	2	AC009218	AC009218 Drosophila melanogaster	
C	49	20.8	83.2	160432	5	AL929072	AL929072 Zebrafish	
C	50	20.8	83.2	162020	9	AC124814	AC124814 Mus musculus	
C	51	20.8	83.2	164236	8	BS000105	BS000105 Pan troglodytes	
C	52	20.8	83.2	165166	5	CR352263	CR352263 Zebrafish	
C	53	20.8	83.2	169614	5	BX511115	BX511115 Zebrafish	
C	54	20.8	83.2	171946	14	AC022359	AC022359 Homo sapiens	
C	55	20.8	83.2	173714	14	AC147748	AC147748 Pan troglodytes	
C	56	20.8	83.2	174804	8	AC087260	AC087260 Homo sapiens	
C	57	20.8	83.2	175403	8	AC093268	AC093268 Homo sapiens	
C	58	20.8	83.2	178637	14	AC163880	AC163880 Bos taurus	
C	59	20.8	83.2	180695	14	AC140661	AC140661 Pan troglodytes	
C	60	20.8	83.2	181927	8	AC114980	AC114980 Homo sapiens	
C	61	20.8	83.2	182193	14	AC147692	AC147692 Gorilla gorilla	
C	62	20.8	83.2	186016	9	AC140458	AC140458 Mus musculus	
C	63	20.8	83.2	189118	9	BX005233	BX005233 Mouse DNA	
C	64	20.8	83.2	192157	2	AC016019	AC016019 Drosophila melanogaster	
C	65	20.8	83.2	193693	14	CR788308	CR788308 Danio rerio	
C	66	20.8	83.2	197507	14	AC140087	AC140087 Pongo pygmaeus	
C	67	20.8	83.2	201360	9	AC116594	AC116594 Mus musculus	
C	68	20.8	83.2	204839	14	AC144882	AC144882 Gorilla gorilla	
C	69	20.8	83.2	206094	14	AC162074	AC162074 Bos taurus	
C	70	20.8	83.2	207362	8	AC007380	AC007380 Homo sapiens	
C	71	20.8	83.2	215018	14	AC139113	AC139113 Pongo pygmaeus	
C	72	20.8	83.2	217329	9	AL611149	AL611149 Mus musculus	
C	73	20.8	83.2	228428	9	AL596456	AL596456 Mouse DNA	
C	74	20.8	83.2	228633	5	BX649600	BX649600 Zebrafish	
C	75	20.8	83.2	228664	14	AC155755	AC155755 Bos taurus	
C	76	20.8	83.2	241203	14	AC152989	AC152989 Bos taurus	
C	77	20.8	83.2	243131	14	AC126866	AC126866 Rattus norvegicus	
C	78	20.8	83.2	250078	2	AE014829	AE014829 Plasmid	
C	79	20.8	83.2	254094	14	AC120220	AC120220 Rattus norvegicus	
C	80	20.8	83.2	263535	2	AE003791	AE003791 Drosophila melanogaster	
C	81	20.8	83.2	349980	6	AX344572	AX344572 Sequence	
C	82	20.4	81.6	349980	763	10	BV069093	BV069093 Caenorhabditis elegans
C	83	20.4	81.6	896	2	AY957636	AY957636 Xenopus laevis	
C	84	20.4	81.6	1187	5	CR762114	CR762114 Xenopus laevis	
C	85	20.4	81.6	27960	2	AF040645	AF040645 Caenorhabditis elegans	
C	86	20.4	81.6	32761	6	AX059464	AX059464 Sequence	
C	87	20.4	81.6	121001	14	AC009328	AC009328 Arabidopsis thaliana	
C	88	20.4	81.6	135686	14	AC140742	AC140742 Rattus norvegicus	
C	89	20.4	81.6	158643	14	AC154144	AC154144 Mus musculus	
C	90	20.4	81.6	160832	14	AC142051	AC142051 Rattus norvegicus	
C	91	20.4	81.6	187545	9	AC127550	AC127550 Mus musculus	

92	20.4	81.6	187755	14	AC006718	AC006718 Caenorhab
C 93	20.4	81.6	202126	9	AC158156	AC158156 Mus muscu
C 94	20.2	80.8	2754	5	XELTGFBS	J05180 X. laevis tr
95	20.2	80.8	3988	5	AJ851646	AJ851646 Gallus ga
96	20.2	80.8	4362	15	AF001415	AF001415 Arabidops
C 97	20.2	80.8	5317	6	AX345511	AX345511 Sequence
C 98	20.2	80.8	5430	6	AX251045	AX251045 Sequence
C 99	20.2	80.8	6880	6	AX251922	AX251922 Sequence
C 100	20.2	80.8	6880	6	AX344318	AX344318 Sequence
C 101	20.2	80.8	6880	6	AX348725	AX348725 Sequence
C 102	20.2	80.8	7455	15	AF520061	AF520061 Candida g
C 103	20.2	80.8	8306	15	AF456838	AF456838 Cryptonec
C 104	20.2	80.8	9369	6	CQ596453	CQ596453 Sequence
C 105	20.2	80.8	10480	6	AX347102	AX347102 Sequence
C 106	20.2	80.8	12366	6	CQ574850	CQ574850 Sequence
C 107	20.2	80.8	35791	14	AC153480	AC153480 Homo sapi
C 108	20.2	80.8	51828	14	AC108345_3	Continuation (4 of
C 109	20.2	80.8	58849	14	AC142514	AC142514 Rattus no
C 110	20.2	80.8	63895	14	AC044880	AC044880 Homo sapi
C 111	20.2	80.8	65993	15	AY459336	AY459336 Oryza sat
C 112	20.2	80.8	68723	14	AC125941_3	Continuation (4 of
C 113	20.2	80.8	69215	14	AC102673	AC102673 Mus muscu
C 114	20.2	80.8	69215	14	AC102673	AC102673 Mus muscu
C 115	20.2	80.8	74687	15	ATPIN13	AL391145 Arabidops
C 116	20.2	80.8	74753	5	CR381565	CR381565 Zebrafish
C 117	20.2	80.8	82436	14	AC130964_3	Continuation (4 of
C 118	20.2	80.8	86588	14	AC128215_3	Continuation (4 of
C 119	20.2	80.8	98382	14	CR936558	CR936558 Danio rer
C 120	20.2	80.8	99688	5	BX539338	BX539338 Zebrafish
C 121	20.2	80.8	100186	14	AC136680	AC136680 Medicago
C 122	20.2	80.8	101391	14	AC134153	AC134153 Rattus no
C 123	20.2	80.8	105305	14	AP007358	AP007358 Lotus cor
C 124	20.2	80.8	105851	14	CT009675	CT009675 Danio rer
C 125	20.2	80.8	106898	14	AC136233	AC136233 Rattus no
C 126	20.2	80.8	110000	1	CR628336_12	Continuation (13 o
C 127	20.2	80.8	110000	1	CR628337_12	Continuation (13 o
C 128	20.2	80.8	110000	14	AC094428_1	Continuation (2 of
C 129	20.2	80.8	110000	14	AC094428_1	Continuation (2 of
C 130	20.2	80.8	110000	14	AC094428_4	Continuation (5 of
C 131	20.2	80.8	110000	14	AC096206_0	AC096206 Rattus no
C 132	20.2	80.8	110000	14	AC097071_0	AC097071 Rattus no
C 133	20.2	80.8	110000	14	AC097071_2	Continuation (3 of
C 134	20.2	80.8	110000	14	AC097441_1	Continuation (2 of
C 135	20.2	80.8	110000	14	AC097441_2	Continuation (3 of
C 136	20.2	80.8	110000	14	AC097441_4	Continuation (5 of
C 137	20.2	80.8	110000	14	AC098269_0	AC098269 Rattus no
C 138	20.2	80.8	110000	14	AC098269_2	Continuation (3 of
C 139	20.2	80.8	110000	14	AC110329_0	AC110329 Rattus no
C 140	20.2	80.8	110000	14	AC110329_2	Continuation (3 of
C 141	20.2	80.8	110000	14	AC110829_0	AC110829 Rattus no
C 142	20.2	80.8	110000	14	AC111557_0	AC111557 Rattus no
C 143	20.2	80.8	110000	14	AC111557_1	Continuation (2 of
C 144	20.2	80.8	110000	14	AC112872_2	Continuation (3 of
C 145	20.2	80.8	110000	14	AC112886_2	Continuation (3 of
C 146	20.2	80.8	110000	14	AC113693_5	Continuation (6 of
C 147	20.2	80.8	110000	14	AC120752_0	AC120752 Rattus no
C 148	20.2	80.8	110000	14	AC120752_1	Continuation (2 of
C 149	20.2	80.8	110000	14	AC120911_3	Continuation (4 of
C 150	20.2	80.8	110000	14	AC128215_0	AC128215 Rattus no

ALIGNMENTS

RESULT 1
 BV550550/c
 LOCUS
 DEFINITION BV550550 719 bp DNA linear STS 09-APR-2005
 sequence tagged site.
 ACCESSION BV550550
 VERSION BV550550.1 GI:62441570
 KEYWORDS
 SOURCE Pan troglodytes troglodytes

ORGANISM Pan troglodytes troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Pan.
 1 (bases 1 to 719)
 Jaffe,D.B.
 TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the
 Human Genome
 COMMENT Unpublished (2005)

Contact: Michael C. Zody
 Broad Institute of MIT and Harvard
 320 Charles Street, Cambridge, MA 02141, USA
 Tel: 6172580933
 Fax: 6172580903
 Email: mczody@broad.mit.edu
 Primer A: No sequence submitted
 Primer B: No sequence submitted
 STS size: 719
 Protocol:

23,021,928 chimpanzee whole genome shotgun reads were aligned to
 the Human genome NCBI
 Build 34 (Hg16, July 2003). Chimp WGS reads were from 9 donors,
 including Clint (Pan
 (Donald, Karlén, Yvonne), 3 Pan
 troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps
 of unknown origin

(don, Unknown Chimp). Common names: Pan troglodytes versus is the
 western chimp and Pan
 troglodytes troglodytes is the central chimp. To be included in
 chimpanzee SNP discovery, a
 read must be at least 500bp in length, at least 50% of its base
 calls must have phred
 score >= 20, at least 30% of its base calls must satisfy
 SNQS(30,25) (single strand NQS, the
 base in question has phred score >= 30, the surrounding 10 bases in
 the read have phred
 score >= 25), and the read must have at least 200 bp SNQS(30,25)
 bases. Reads not uniquely
 placed in the genome and read pairs whose two ends were not
 consistently placed were
 discarded. After above filtering, NQS(30,25) standard was applied
 to all pairs of
 overlapping reads to call NQS bases and SNPs. Alignments (between
 two reads) with less
 than 100 NQS bases or with SNP rate > 0.01 were discarded. To
 exclude alignment between two
 copies of a single read, comparisons between two reads that share
 95% of their genome
 alignments (>=95% bases of read A and >=95% bases of read B were
 placed at the same locus
 of human genome) were discarded.

FEATURES

source
 1..719
 /organism="Pan troglodytes troglodytes"
 /mol_type="genomic DNA"
 /sub_species="troglodytes"
 /db_xref="taxon:37011"
 /clone_lib="Masuku"
 <1..>719

ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 719;
 Best Local Similarity 96.0%; Pred. No. 4.8e+02;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 AAAAAAAAACTTCATCATTTAAAC 25
 Db 593 AAAAAAAAACTTCATCATTTAAAC 569

RESULT 2
AF131217
LOCUS
DEFINITION Homo sapiens chromosome 21 clone PAC G1664 map 21q22.1, complete sequence.
ACCESSION AF131217
VERSION AF131217.3 GI:14475923
KEYWORDS HMG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 130530)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieselmann,L., Dagand,E., Haaf,T., Wehrmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M. laure.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
10830953

TITLE
JOURNAL PUBMED
REFERENCE 2 (bases 1 to 130530)
AUTHORS Reichwald,K., Delabar,J., Orti,R., Dagand,E., Baumgart,C., Nordstiek,G., Menzel,U., Yaspo,M.L. and Rosenthal,A.
Direct Submission
JOURNAL Submitted (25-FEB-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
3 (bases 1 to 130530)
Reichwald,K., Delabar,J., Orti,R., Dagand,E., Baumgart,C., Nordstiek,G., Menzel,U., Yaspo,M.L. and Rosenthal,A.
Direct Submission
JOURNAL Submitted (27-AUG-1999) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
Sequence update by submitter
REMARK 4 (bases 1 to 130530)
REFERENCE Reichwald,K., Delabar,J., Orti,R., Dagand,E., Baumgart,C., Nordstiek,G., Menzel,U., Yaspo,M.-L. and Rosenthal,A.
JOURNAL Direct Submission
TITLE Submitted (16-JUN-2001) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Jun 16, 2001 this sequence version replaced gi:5787974.
FEATURES
SOURCE
1. 130530
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="PAC G1664"

ORIGIN
Query Match 93.6%; Score 23.4; DB 8; Length 130530;
Best Local Similarity 96.0%; Pred. No. 84;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 56824 AAAAAAAAAATTCATCATTTAAAC 25
1 AAAAAAAAAACCTCATCATTTAAAC 25
AAAAAAAAAAAAAAAAATTCATCATTTAAAC 56848

RESULT 3
HS21C047
LOCUS
DEFINITION Homo sapiens chromosome 21 segment HS21C047.

ACCESSION AL163247 AP001702 BA000005
VERSION AL163247.2 GI:7717303
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E., Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibusaki,K., Kawasaki,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Rieselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Yaspo,M.L.
Direct Submission
JOURNAL Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GSF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of * RIKEN Genomic Sciences Center, Human Genome Research Group, Sagamihara 228-8555, Japan, * e.mail: sasaki@gsc.riken.go.jp * URL: http://hgp.gsc.riken.go.jp/ and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany, * e.mail: gscj-submit@genome.imb-jena.de * URL: http://genome.imb-jena.de/ and
* Keio University School of Medicine, Dept. of Molecular Biology, * Tokyo 160-8582, Japan, * e.mail: shimizu@cmb-med.keio.ac.jp * URL: http://adenine.dmb.med.keio.ac.jp/ and
* GSF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de * URL: http://genome.gbf.de/ and
* Max-Planck Institute for Molecular Genetics, * Inestrasse 73, D-14195 Berlin, Germany, * e.mail: info-chr21@molgen.mpg.de * URL: http://chr21.tz-berlin.mpg.de/.
Location/Qualifiers
1. 340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
<1. 81759
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="P775E24, 5' partial"
/clone.lib="RP11.3-5 PAC library"
/note="Accession No. AL035534"
29024. 159555
/organism="Homo sapiens"
/mol_type="genomic DNA"

source

```

/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="P64G16"
/clone_id="RPC11,3-5 PAC library"
/note="Accession No. AF131217"
118472. .265213
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="P80122"
/clone_id="RPC11,3-5 PAC library"
/note="Accession No. AF165147"
26216. .340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="P38822, 3' partial"
/note="Accession No. AF129078"
293286. .>340000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="P39N21, 3' partial"
/clone_id="RPC11,3-5 PAC library"
/note="Accession No. AF207955"
1825. .2081
/note="HERV L"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
2084. .2503
/clone="MLT2A"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
3763. .3969
/clone="MLT1J1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
4285. .4339
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
4935. .5015
/note="LTR33A"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
6389. .6517
/note="LIMEC"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
6518. .6548
/note="(CAAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
6549. .8334
/note="LIMEC"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
8335. .8953
/note="MER76"
/rpt_family="LTR"
/rpt_type=DISPERSED
8978. .9298
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED

```

```

repeat_region 9299. .9590
/note="LIMEC"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
9769. .9816
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
9948. .10007
/note="MER76"
/rpt_family="LTR"
/rpt_type=DISPERSED
10013. .10046
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
10099. .10141
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
10142. .10957
/note="LIM4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
10971. .11080
/note="LTR40b"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
11963. .12322
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
12325. .13861
/note="THE1C-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
13862. .14248
/note="THE1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(14724. .14848)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
14904. .14966
/note="(TATAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
15232. .15620
/note="MSTD"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED

```

Query Match 93.6%; Score 23.4; DB 8; Length 340000;
 Best Local Similarity 96.0%; Pred. NO. 61;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25
Db 85847 AAAAAAAAACTTCATCATTTAAAC 85871

RESULT 4
AC114719 146608 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-122124, WORKING DRAFT SEQUENCE, 3
DEFINITION
unordered pieces.
AC114719
AC114719.4 GI:25072889
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLOR.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 146608)

TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

Munry, D. Marie, Metzger, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bivawala, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Diaper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Far, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, N., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangun, A., Mangum, B., Manua, P., Martin, K., Martin, R., Martinez, E., Manthey, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Narkewicz, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokoileme, O., Okunolu, G., Olarnpungoon, A., Pal, S., Parke, K., Pasternak, S., Paul, H., Perez, A., Perez, J., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, B., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Reijer, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaje, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabot, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wiczzyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., and Gibbs, R.A.

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (11-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 146608)
Rat Genome Sequencing Consortium.

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23267305.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GT1Y
Center clone name: CH230-122124
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 14314 bases at least Q40
Consensus quality: 144271 bases at least Q30
Consensus quality: 144860 bases at least Q20
Estimated insert size: 14193; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preselected.

1 104033: contig of 104033 bp in length
* 104034 104133: gap of unknown length
* 104134 140217: contig of 36084 bp in length
* 140218 140317: gap of unknown length
* 140318 146608: contig of 6291 bp in length.
Location/Qualifiers
1. 146608

FEATURES
source

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-122124"
122. 975
misc_feature
122. 975
site:
end_sequence="BH269675"
complement(99839..100712)
/note="clone boundary
clone_end:Sp6
site:
end_sequence="BH269676"
101422..102445
/note="wgs_end_extension
clone_end:Sp6"
103004..104033
/note="wgs_end_extension
clone_end:Sp6"
104034..104133
/estimated_length=unknown
140218..140317
/estimated_length=unknown
140318..141348
/note="wgs_end_extension
clone_end:Sp6"

ORIGIN

Query Match 89.6%; Score 22.4; DB 14; Length 146608;
Best Local Similarity 95.8%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0;
1 AAAAAAAAACTTCATCATTTAA 24

Db 36123 AAAAAAAAAACCTCATCTTTAAA 36146

RESULT 5 AC098628

DEFINITION Rattus norvegicus clone CH230-108020, *** SEQUENCING IN PROGRESS
LOCUS *** 2 unordered pieces.
ACCESSION AC098628 237844 bp DNA linear HTG 10-MAY-2003
VERSION AC098628.7 GI:30520732
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Eutheria; Euarctomolgidae; Glires; Rodentia;
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Sciuromorphi; Muridae; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 237844)
AUTHORS Murthy, D., Marie, Metzger, M., Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Albrooks, S., Amin, A., Arguiano, D.,
Anyalbechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davi, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denon, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Dreber, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Geregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevarra, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Henningsen, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howell, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovle, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Mlisa, J., Mlisa, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newcom, N., Nguyen, N., Norris, S.,
Nwokedi, O., Okwona, G., Olariu, S., Olatunji, A., Olatunji, S.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L. L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X. Z., Sorrell, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,
Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wiczysk, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

REFERENCE AUTHORS TITLE JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237844)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:25008275.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHXB
Center clone name: CH230-108020
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 198885 bases at least Q40
Consensus quality: 203692 bases at least Q30
Consensus quality: 206750 bases at least Q20
Estimated insert size: 212383; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 232734: contig of 232734 bp in length
* 232735 232834: gap of unknown length
* 232835 237844: contig of 5010 bp in length.
Location/Qualifiers
1..237844
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-108020"
1..2086
/note="wgs contig"
21283..23767
/note="wgs contig"
101690..104702
/note="wgs contig"
108403..110059
/note="wgs contig"
232735..232834
/estimated_length=unknown
ORIGIN
Query Match 89.6%; Score 22.4; DB 14; Length 237844;
Best Local Similarity 95.8%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
by 1 AAAAAAAAAACCTCATCTTTAAA 24

```

Db      85423  AAAAAAAAAACCTTCATCTTTAA 85446
|||||
RESULT 6
LOCUS   AX397279          480 bp    DNA          linear    PAT 18-MAY-2002
DEFINITION Sequence 1494 from Patent WO0212328.
ACCESSION AX397279
VERSION   AX397279.1  GI:21068026
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homiidae; Homo.
REFERENCE
AUTHORS 1 King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE    Compositions and methods for the therapy and diagnosis of colon
JOURNAL  Cancer
PATENT   Patent: WO 0212328-A 1494 14-FEB-2002;
FEATURES
SOURCE   1. 480
           /organism="Homo sapiens"
           /mol_type="unassigned DNA"
           /db_xref="taxon:9606"

ORIGIN
Query Match      87.2%; Score 21.8; DB 6; Length 480;
Best Local Similarity 92.0%; Pred. No. 2e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCTTTAAAC 25
Db 188 AAAAAAAAAACCTTCACATTTATAC 164
|||||
RESULT 7
LOCUS   BV522543          904 bp    DNA          linear    STS 08-APR-2005
DEFINITION GS91P68042RG8.T0 Clint Pan troglodytes versus STS genomic, sequence
tagged site.
ACCESSION BV522543
VERSION   BV522543.1  GI:62400313
KEYWORDS
SOURCE   Pan troglodytes versus
ORGANISM Pan troglodytes
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homiidae; Pan.
REFERENCE 1 (bases 1 to 904)
AUTHORS  Mikkelson, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and
           Jaffe, D.B.
TITLE    Initial Sequence of the Chimpanzee Genome and Comparison with the
JOURNAL  Human Genome
COMMENT   Unpublished (2005)
           Contact: Michael C. Zody
           Broad Institute of MIT and Harvard
           320 Charles Street, Cambridge, MA 02141, USA
           Tel: 6172580933
           Fax: 6172580903
           Email: mczody@broad.mit.edu
           Primer A: No sequence submitted
           Primer B: No sequence submitted
           STS size: 904
           Protocol:
           23,021,928 chimpanzee whole genome shotgun reads were aligned to
           the Human genome NCBI
           Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,
           including Clint (Pan
           troglodytes versus), 3 other Pan troglodytes versus chimps

```

```

(Donald, Karlén, Yvonne), 3 Pan
troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps
of unknown origin
(Gon, Unknown Chimp). Common names: Pan troglodytes versus is the
western chimp and Pan
troglodytes troglodytes is the central chimp. To be included in
chimpanzee SNP discovery, a
read must be at least 500bp in length, at least 50% of its base
calls must have Phred
score >= 20, at least 30% of its base calls must satisfy
SNOS(30,25) (single strand NOS, the
base in question has Phred score >= 30, the surrounding 10 bases in
the read have Phred
score >= 25), and the read must have at least 200 bp SNOS(30,25)
bases. Reads not uniquely
placed in the genome and read pairs whose two ends were not
consistently placed were
discarded. After above filtering, NOS(30,25) standard was applied
to all pairs of
overlapping reads to call NOS bases and SNPs. Alignments (between
two reads) with less
than 100 NOS bases or with SNP rate > 0.01 were discarded. To
exclude alignment between two
copies of a single read, comparisons between two reads that share
95% of their genome
alignments (>=95% bases of read A and >=95% bases of read B were
placed at the same locus
of human genome) were discarded.
location/Qualifiers
1. 904
/organism="Pan troglodytes versus"
/mol_type="genomic DNA"
/sub_species="versus"
/db_xref="taxon:37012"
/clone_1ib="Clint"
<1..>904

ORIGIN
STS
Query Match      87.2%; Score 21.8; DB 10; Length 904;
Best Local Similarity 92.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCTTTAAAC 25
Db 471 AAAAAAAAAATCTTCACCATTTAAAC 447
|||||
RESULT 8
LOCUS   AC111656_4
WPCOMMENT
Sequence split into 5 fragments LOCUS AC111656 Accession AC111656
Fragment Name      Begin      End
AC111656_0         1      110000
AC111656_1        100001    210000
AC111656_2        200001    310000
AC111656_3        300001    410000
AC111656_4        400001    430740
Continuation (5 of 5) of AC111656 from base 400001 (AC111656 Ratius norvegicus clone CH

Query Match      87.2%; Score 21.8; DB 14; Length 30740;
Best Local Similarity 92.0%; Pred. No. 4.9e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCTTTAAAC 25
Db 5482 AAAAAAAAAATCTTCATCTTTAAAC 5506
|||||
RESULT 9
LOCUS   AC006129          38531 bp    DNA          linear    PRI 17-DEC-1998
DEFINITION Homo sapiens chromosome 19, cosmid F6697, complete sequence.
ACCESSION AC006129

```

```

VERSION      AC006129.1  GI:3970933
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homnidae; Homo.
REFERENCE    1 (bases 1 to 38531)
AUTHORS      Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V.,
              Burkhardt-Schultz,K., Gordon,L., Dias,J., Brower,A., Ramirez,M.,
              Stilhagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A.,
              Ganev,J., Danganan,L., Erler,A., Christensen,M., Georgescu,A.,
              Avila,J., Liu,S., Attix,C., Andreise,T., Frankheim,M.,
              Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R.,
              Thomas,P., Quan,G., Krommiller,B., Arellano,A., Sanders,C., Ow,D.,
              Notian,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
              Sequence analysis of a 2.5 Mb region in 19q13.2 containing a
              clustered CEA/PSG gene family
              Unpublished
              2 (bases 1 to 38531)
              Lamerdin,J.E.
              Direct Submission
              Submitted (05-DEC-1998) Joint Genome Institute, Lawrence Livermore
              National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
              3 (bases 1 to 38531)
              Lamerdin,J.E.
              Direct Submission
              Submitted (17-DEC-1998) Joint Genome Institute, Lawrence Livermore
              National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
              Map and sequence are oriented from q centromere to telomere. Cosmid
              F6697 overlaps cosmid F20649 (AC005795) to the left from bases 1 to
              5,380 of this accession and overlaps cosmid R29124 (AC005626) to
              the right from bases 32,348 to 38,531. Additional map and sequence
              information may be obtained at:
              http://www-bio.lnl.gov/brp/genome/genome.html.
              Location/Qualifiers
              1..38531
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="19"
              /map="19q13.2 between BCKDA and D19S217"
              /clone="F6697"
              /cell_line="UV5HL9-5B"
              /clone_lib="L19NC02 F chromosome 19-specific cosmid
              library"
              /note="Cosmid library constructed at LNL from flow-sorted
              chromosomes from hybrid UV5HL9-5B, which carries
              chromosome 19 as its only human chromosome."
              652..736
              /rpt_family="GC_rich"
              /note="predicted exon, program: grail2exons_human_1.3,
              frame: 1, quality: excellent, score: 79.000"
              /note="predicted exon, program: grail2exons_human_1.3,
              complement(1279..2501)
              frame: 1, quality: good, score: 54.000"
              3460..3550
              /note="predicted exon, program: grail2exons_human_1.3,
              frame: 1, quality: good, score: 52.000"
              4289..4493
              /rpt_family="L1PA7"
              4504..5036
              /rpt_family="L1"
              5038..5322
              /rpt_family="AluSc"
              5327..6205
              /rpt_family="L1"
              /note="complement(6635..7189)
              /rpt_family="LTR10C"
              repeat_region complement(7192..7483)
              /rpt_family="AluX"
              9072..9139
              repeat_region

```

```

              /rpt_family="LINE2"
              complement(10956..11626)
              /rpt_family="MER41_internal"
              complement(11639..12346)
              /rpt_family="L1PB3"
              complement(12347..12598)
              /rpt_family="LTR4"
              complement(12702..12768)
              /note="predicted exon, program: grail2exons_human_1.3,
              frame: 0, quality: good, score: 61.000"
              13040..13586
              /rpt_family="L1"
              13539..14092
              /rpt_family="L1"
              complement(14252..14284)
              /rpt_family="AT_rich"
              14348..14418
              /note="DDB similarity to AA832417 oc99d03.g1 NCI CGAP_GCB1
              Homo sapiens cDNA clone IMAGE:1357829 3' similar to
              contains LTR1.t2 repetitive element ; (395..326), 100%
              identity."
              complement(14563..14615)
              /rpt_family="AT_rich"
              15038..15296
              /rpt_family="LTR1"
              15236..15539
              /note="DDB similarity to AA832417 oc99d03.g1 NCI CGAP_GCB1
              Homo sapiens cDNA clone IMAGE:1357829 3' similar to
              contains LTR1.t2 repetitive element ; (325..22), 99%
              identity."
              15371..15870
              /rpt_family="LTR1"
              17668..18344
              /rpt_family="HERV17"
              18369..18453
              /rpt_family="HERV17"
              18506..19161
              /rpt_family="LTR1"
              complement(19368..20662)
              /rpt_family="MER57_internal"
              complement(20696..20758)
              /rpt_family="AT_rich"
              complement(21007..21201)
              /rpt_family="MLT1D"
              complement(21727..21896)
              /rpt_family="FRAM"
              complement(21900..21926)
              /rpt_family="AT_rich"
              21934..22163
              /note="predicted exon, program: grail2exons_human_1.3,
              frame: 2, quality: good, score: 58.000"
              23213..24242
              /rpt_family="AT_rich"
              24503..24587
              /rpt_family="LTR5"
              /note="predicted exon, program: grail2exons_human_1.3,
              frame: 0, quality: good, score: 66.000"
              25380..25432
              /rpt_family="(CA)n"
              25552..25623
              /rpt_family="AluS"
              26085..26256
              /rpt_family="MER4_internal"
              26237..26418
              /note="predicted exon, program: grail2exons_human_1.3,
              frame: 0, quality: excellent, score: 79.000"
              26421..26618
              /note="predicted exon, program: grail2exons_human_1.3,
              frame: 2, quality: good, score: 52.000"
              /rpt_family="MER41_internal"
              complement(28204..28329)
              repeat_region

```

```

repeat_region      /rpt_family="(CAAAA)n"
complement(28330..28501)
misc_feature        /rpt_family="FRAM"
28659..28702
/notes="predicted exon, program: grai12exons_human_1.3,
frame: 1, quality: excellent, score: 95.000"
repeat_region      /rpt_family="AluSx"
29130..29432
/notes="predicted exon, program: grai12exons_human_1.3,
frame: 1, quality: excellent, score: 95.000"
repeat_region      /rpt_family="(MER4_internal"
complement(29802..30100)
/rpt_family="AluSg"
repeat_region      /rpt_family="(MER4_internal"
complement(30562..31225)
/rpt_family="PAB1_A"
repeat_region      /rpt_family="(MER4_internal"
complement(31467..31765)
/rpt_family="AluSx"
repeat_region      /rpt_family="(MER4_internal"
complement(31766..32046)
/rpt_family="AluSx"
repeat_region      /rpt_family="(MER4_internal"
complement(32047..33608)
/rpt_family="(MER41_internal"
complement(33842..33969)
/rpt_family="POLY_A"
misc_feature        /rpt_family="POLY_A"
34603..34728
/notes="predicted exon, program: grai12exons_human_1.3,
frame: 0, quality: excellent, score: 76.000"
misc_feature        /rpt_family="(MER57_internal"
34791..34879
/notes="predicted exon, program: grai12exons_human_1.3,
frame: 2, quality: good, score: 69.000"
repeat_region      /rpt_family="(L1PA16"
complement(34942..35001)
/rpt_family="(L1PA16"
repeat_region      /rpt_family="(MER57_internal"
complement(35044..35256)
/rpt_family="(MER57_internal"
complement(36501..36635)
/rpt_family="MIR"
repeat_region      /rpt_family="AluSg"
37010..37309
/rpt_family="(TAAA)n"
repeat_region      /rpt_family="(TAAA)n"
38174..38256
/rpt_family="(TAAA)n"
repeat_region      /rpt_family="(TAAA)n"
38464..38531
/rpt_family="AluSp"
repeat_region      /rpt_family="AluSp"
38464..38531

```

ORIGIN

```

Query Match      87.2% Score 21.8; DB 8; Length 38511;
Best Local Similarity 92.0%; Pred. No. 4.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 37305 AAAAAAAAAACCTTCATCATTTATAC 37329

```

```

RESULT 10
AC005626      39233 bp DNA linear PRI 06-SEP-1998
LOCUS Homo sapiens chromosome 19, cosmid R29124, complete sequence.
DEFINITION AC005626
AC005626.1 GI:3549151
VERSION AC005626.1 GI:3549151
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia: Eutheria; Barchonotoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 39233)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,J., Kyle,A., Ramirez,M., Stiilwagen,S.,
Plan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garne,J.,
Dangnan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,T.,
Liu,S., Actix,C., Andreise,T., Frankheim,W., Amico-Keller,G.,
Coeffield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Krommiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of a 2.5 Mb region in 19q13.2 containing a
clustered CBA/PSG gene family
unpublished

```

```

REFERENCE 2 (bases 1 to 39233)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Map and sequence oriented from q centromere to telomere. Cosmid
R29124 overlaps cosmid F6697 to the left by at least 6 kb and is
expected to overlap cosmid R32065 to the right by approximately 5
kb. Additional chr 19 map and sequence information are available
at: http://www-bio.lnl.gov/genome/genome.html.
FEATURES
source
1..39233
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2 between BCKDHA and D19S217"
/clone="R29124"
/cell_line="SHL2-B"
/clone_1lb="DL19NC03 R chromosome 19-specific cosmid
library"
/notes="Cosmid library constructed at LNL from flow-sorted
chromosomes from hybrid SHL2-B, which carries chromosome
19 as its only human chromosome."
complement(1..1261)
/rpt_family="(MER41_internal"
complement(1595..1622)
/rpt_family="POLY_A"
2256..2381
/notes="predicted exon, program: grai12exons_human_1.3,
frame: 2, quality: excellent, score: 76.000"
2444..2532
/notes="predicted exon, program: grai12exons_human_1.3,
frame: 1, quality: good, score: 69.000"
complement(2595..2654)
/rpt_family="(L1PA16"
complement(4154..4288)
/rpt_family="MIR"
4663..4962
/rpt_family="AluSg"
5827..5909
/rpt_family="(TAAA)n"
6124..6302
/rpt_family="FAM"
6310..6502
/rpt_family="L1"
complement(6517..6776)
/rpt_family="AluJo"
6778..7474
/rpt_family="L1"
complement(7477..7593)
/rpt_family="(FLAM_A"
7758..7918
/rpt_family="L1"
8260..8560
/rpt_family="AluSx"
8561..8644
/rpt_family="(GAAA)n"
8755..8894
/rpt_family="(L1ME3"
complement(9024..9323)
/rpt_family="AluY"
9349..9507
/notes="BLASTN similarity to cp923698.cta (2..160); match:
0.99, score: 4.6e-58; database searched: Sanger CPG; bases
98 to 257 (SL to QR)"
complement(9349..9504)
/notes="BLASTN similarity to cp923698.cta (15..170);
match: 0.99, score: 7.7e-57; database searched: Sanger
CPG; bases 78 to 247 (SL to SR)"
complement(9916..10068)
/notes="BLASTN similarity to M11167 (4862..5014); match:
0.74, score: 5.4e-139; database searched: nt; Human 285

```

```

misc_feature      ribosomal RNA gene."
                  complement (10138..10222)
                  /note="BLASTN similarity to M11167 (4652..4736); match:
                  0.75, score: 5.4e-139; database searched: nt; Human 285
misc_feature      ribosomal RNA gene."
                  complement (10252..10371)
                  /note="BLASTN similarity to M11167 (4501..4620); match:
                  0.85, score: 5.4e-139; database searched: nt; Human 285
misc_feature      ribosomal RNA gene."
                  complement (10419..10613)
                  /note="BLASTN similarity to M11167 (4248..4442); match:
                  0.82, score: 5.4e-139; database searched: nt; Human 285
repeat_region     ribosomal RNA gene."
                  complement (10621..10652)
                  /rpt_family="TAAAN"
misc_feature      complement (10703..10728)
                  /note="BLASTN similarity to M11167 (4161..4186); match:
                  0.88, score: 5.4e-139; database searched: nt; Human 285
misc_feature      ribosomal RNA gene."
                  complement (10858..10888)
                  /note="BLASTN similarity to M11167 (3947..3977); match:
                  0.8, score: 5.4e-139; database searched: nt; Human 285
misc_feature      ribosomal RNA gene."
                  complement (11039..11097)
                  /note="BLASTN similarity to M11167 (3723..3781); match:
                  0.83, score: 5.4e-139; database searched: nt; Human 285
misc_feature      ribosomal RNA gene."
                  complement (11160..11735)
                  /note="BLASTN similarity to M11167 (2803..2878); match:
                  0.82, score: 5.4e-139; database searched: nt; Human 285
misc_feature      ribosomal RNA gene."
                  complement (11814..11834)
                  /rpt_family="GC rich"
repeat_region     complement (11853..11920)
                  /rpt_family="AluIo/FRAM"
repeat_region     complement (12701..13337)
                  /rpt_family="L1"
repeat_region     complement (13751..13625)
                  /rpt_family="L1"
repeat_region     complement (14415..15340)
                  /rpt_family="L1"
repeat_region     complement (15486..15428)
                  /rpt_family="L1MA6"
repeat_region     complement (15493..15776)
                  /rpt_family="AluIo"
misc_feature      complement (16303..16256)
                  /note="predicted exon, program: graal2exons_human_1.3,
                  frame: 1, quality: excellent, score: 89.000"
repeat_region     complement (16664..16583)
                  /rpt_family="MER22"
repeat_region     complement (17326..17471)
                  /rpt_family="MER20"
misc_feature      complement (19263..19128)
                  /note="predicted exon, program: graal2exons_human_1.3,
                  frame: 2, quality: good, score: 64.000"
repeat_region     complement (20153..20153)
                  /rpt_family="L1MB5"
misc_feature      complement (22553..22616,23478..23871,30641..30722,31722..31784,
                  32387..32474,32743..32805)
                  /note="CGMa-like pseudogene"
CDS               join(22553..22616,23478..23871,30641..30722,31722..31784,
                  32387..32474,32743..32805)
                  /note="Hypothetical human CDM2-like protein"
                  /codon_start=1
                  /evidence=not experimental
                  /product="R29124_1"
                  /protein_id="AAC34569.1"
                  /db_xref="GI:3549152"
misc_feature      translation="MGPPSACPHRECI PMOGLLTASILTFMNAFTTAMLFASAPFE
                  VAEGENHLASVYLPENLISYGMKGTVEPQILAAVIDTHVTPGPAYSGRETTIS
                  PSGDLHPQNTLEDGYTYLTQVTRNSQIEQASHIRYSEVAOPSIOASSTYLERK
                  SVVLCTNNNTGTSFQWIFNNQRLQVTKMKLSWNHMLTIDPLKQEDAGEYQCEVSN
                  EVSSNRSDPLKLTIVCE"
misc_feature      22670..22696

```

```

                  /note="BLASTX similarity to 437784 (36..44); match: 0.88,
                  score: 3.8e-15; database searched: nr; (Z21818)
                  carcinoembryonic antigen [Homo sapiens]"
                  complement (23008..23329)
                  /rpt_family="L1MB2"
misc_feature      25583..25678
                  /note="BLASTX similarity to (307..338); match: 0.43,
                  score: 1.1e-68; database searched: nr; transmembrane
                  carcinoembryonic antigen 3 precursor - human >gi||550031
                  (571326) Bgpc-biliary glycoprotein adhesion molecule"
misc_feature      25632..25907
                  /note="predicted exon, program: graal2exons_human_1.3,
                  frame: 1, quality: excellent, score: 93.000"
repeat_region     complement (26021..26257)
                  /rpt_family="WIR"
repeat_region     complement (27303..26556)
                  /rpt_family="MER21B"
repeat_region     complement (27314..27821)
                  /rpt_family="MER9"
repeat_region     complement (27822..28149)
                  /rpt_family="THER1B"
repeat_region     complement (28163..29757)
                  /rpt_family="THER1B-internal"
repeat_region     complement (29760..30108)
                  /rpt_family="THER1B"
misc_feature      30643..30723
                  /note="BLASTX similarity to (155..181); match: 0.77,
                  score: 3.7e-18; database searched: nr; non-specific
                  cross-reacting antigen W236 precursor - human
                  >gnl|PIID1015026 (D90276) CCM7 [Homo sapiens]"
                  30869..30918
                  /rpt_family="WIR"
repeat_region     Query Match
                  Best Local Similarity 87.2%; Score 21.8; DB 8; Length 39233;
                  Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
                  Db 4958 AAAAAAAAAACCTTCATCATTTAAAC 25
OY 1 AAAAAAAAAACCTTCATCATTTAAAC 25
WPCOMMENT
Sequence split into 5 fragments LOCUS AC111656 Accession AC111656
Fragment Name Begin End
AC111656_0 1 110000
AC111656_1 100001 210000
AC111656_2 200001 310000
AC111656_3 300001 410000
AC111656_4 400001 430740
Continuation 74 of 5 of AC111656 from base 300001 (AC111656 Ratius norvegicus clone CH
Query Match 87.2%; Score 21.8; DB 14; Length 110000;
Best Local Similarity 92.0%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 105482 AAAAAAAAAATTCATCATTTAAAC 105506
RESULT 12
AC128588_0
WPCOMMENT
Sequence split into 5 fragments LOCUS AC128588 Accession AC128588
Fragment Name Begin End
AC128588_0 1 110000
AC128588_1 100001 210000
AC128588_2 200001 310000
AC128588_3 300001 410000
AC128588_4 400001 461536

```


LOCUS	461536 bp	DNA	linear	HTG 12-OCT-2002
DEFINITION	Rattus norvegicus clone CH230-233K11. **** SEQUENCING IN PROGRESS			
ACCESSION	AC128598			
VERSION	AC128598.2 GI:2307860			
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.			
	1 (bases 1 to 461536)			
AUTHORS	Allen, D., Marie, Metzker, M. Lee, Abramzon, S., Adams, C., Alder, J.,			
	Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Ayalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Bernhamed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, C., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesaar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Andrade, C., Dedrich, D., Delgado, O., Denson, S., Derriso, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P., Fraser, C. M., Gabisi, A., Gante, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kwis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, Y., London, P., Longacre, S., Lopez, J., Loreneuhewa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maneswarthi, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangun, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Mnja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankevici, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokeleneh, O., Okunonu, G., Olampunsgoon, A., Pal, S., Parks, K., Pasternack, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L., Pu, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Kelly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sibson, I., Sitter, C. D., Snaib, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorcelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Swatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemari, K., Vaas, R., Vera, V., Villasaana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Wernstedt, G. A., and Gibbs, R. A.			
TITLE	Direct Submission			
	2 (bases 1 to 461536)			
JOURNAL	Unpublished			
	Worley, K. C.			
AUTHORS	Direct Submission			
	Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
REFERENCE	3 (bases 1 to 461536)			
	Rat Genome Sequencing Consortium.			
TITLE	Direct Submission			
	Submitted (12-OCT-2002) Human Genome Sequencing Center, Department			

COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 12, 2002 this sequence version replaced gi:21909369.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Center project name: GYNS
Center clone name: CH230-233K11

----- Summary Statistics -----

Assembly program: Phrap; version 0.990329
Consensus quality: 249289 bases at least Q40
Consensus quality: 256050 bases at least Q30
Consensus quality: 259296 bases at least Q20
Estimated insert size: 279061; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 11520: contig of 11520 bp in length
* 11521 11620: gap of unknown length
* 11621 22497: contig of 10877 bp in length
* 22498 22597: gap of unknown length
* 22598 453608: contig of 431011 bp in length
* 453609 453708: gap of unknown length
* 453709 455036: contig of 1328 bp in length
* 455037 455136: gap of unknown length
* 455137 461536: contig of 6400 bp in length.

Location/Qualifiers

1. 461536

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-233K11"
complement(3940..4743)
/note="clone_boundary
clone_end:Sp6
site:EcorI
end_sequence:RMBOG6TV"
9481..11520
/note="wgs_contig"
11521..11620
/estimated_length=unknown
11621..13033
/note="wgs_contig"
14127..15525
/note="wgs_contig"
20336..22497
/note="wgs_contig"
22498..22597

FEATURES

SOURCE

misc_feature

misc_feature

misc_feature

misc_feature

gap

```
misc_feature      /estimated_length=unknown
                  54628..55534
                  /note="clone_boundary
                  clone_end:T7
                  site:ECORI
                  end_sequence:RMBQO66TJ"
misc_feature      69564..71337
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature      96528..98369
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature      11952..121084
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature      146503..148458
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature      183429..184920
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature      452267..453608
                  /note="wgs_end_extension
                  clone_end:T7"
misc_feature      453609..453708
                  /note="wgs_end_extension
                  clone_end:T7"
gap              /estimated_length=unknown
                  455037..455136
gap              /estimated_length=unknown

ORIGIN
Query Match      87.2%; Score 21.8; DB 14; Length 110000;
Best Local Similarity 92.0%; Pred.No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY              1 AAAAAAAAACTTCATCATTTAAAC 25
Db              21638 AAAAAAAAAATTCATCATTTAAAC 21662

RESULT 13
CR555291_0
WPCOMMENT
Sequence split into 5 fragments LOCUS CR555291 Accession CR555291
Fragment Name      Begin      End
CR555291_0          1      110000
CR555291_1          100001  210000
CR555291_2          200001  310000
CR555291_3          300001  410000
CR555291_4          400001  410877
LOCUS CR555291 410877 bp DNA linear HTG 11-OCT-2004
DEFINITION Danio rerio clone DKEX-58J6, ** SEQUENCING IN PROGRESS ***, 37
          unordered pieces.
ACCESSION CR555291
VERSION CR555291.4 GI:52313314
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
          1 (bases 1 to 410877)
McLay, K.
Direct Submission
Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
clonerequests@sanger.ac.uk
zfish-help@sanger.ac.uk
On Sep 17, 2004 this sequence version replaced gi:50724921.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
```

```
Center project name: ZK58J8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 399430 bases at least Q40
Consensus quality: 402692 bases at least Q30
Consensus quality: 407277; sum-of-contigs
Insert size: 200961; 2.9% error; agarose-fp
Quality coverage: 4.75x in Q20 bases; sum-of-contigs Quality
coverage: 9.83x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 5913: contig of 5913 bp in length
* 5914 6013: gap of 100 bp
* 6014 8737: contig of 2724 bp in length
* 8738 8837: gap of 100 bp
* 8838 12348: contig of 3511 bp in length
* 12349 12448: gap of 100 bp
* 12449 34866: contig of 22418 bp in length
* 34867 34966: gap of 100 bp
* 34967 50012: contig of 15046 bp in length
* 50013 50112: gap of 100 bp
* 50113 52494: contig of 2382 bp in length
* 52495 52594: gap of 100 bp
* 52595 55884: contig of 3290 bp in length
* 55885 55984: gap of 100 bp
* 55985 59774: contig of 3790 bp in length
* 59775 59874: gap of 100 bp
* 59875 71704: contig of 11830 bp in length
* 71705 71805: gap of 100 bp
* 71805 90091: contig of 18287 bp in length
* 90092 90191: gap of 100 bp
* 90192 94302: contig of 4111 bp in length
* 94303 94402: gap of 100 bp
* 94403 104620: contig of 10218 bp in length
* 104621 104720: gap of 100 bp
* 104721 110745: contig of 6025 bp in length
* 110746 110845: gap of 100 bp
* 110846 113873: contig of 3028 bp in length
* 113874 113973: gap of 100 bp
* 113974 118974: contig of 5001 bp in length
* 118975 119074: gap of 100 bp
* 119075 139540: contig of 20466 bp in length
* 139541 139640: gap of 100 bp
* 139641 146697: contig of 7057 bp in length
* 146698 146797: gap of 100 bp
* 146798 149204: contig of 2407 bp in length
* 149205 149304: gap of 100 bp
* 149305 158380: contig of 9076 bp in length
* 158381 158480: gap of 100 bp
* 158481 176802: contig of 18322 bp in length
* 176803 176902: gap of 100 bp
* 176903 179349: contig of 2447 bp in length
* 179350 179449: gap of 100 bp
* 179450 183727: contig of 4278 bp in length
* 183728 183827: gap of 100 bp
* 183828 186517: contig of 2690 bp in length
* 186518 186617: gap of 100 bp
* 186618 195486: contig of 8871 bp in length
* 195487 195588: gap of 100 bp
* 195589 214558: contig of 18970 bp in length
* 214559 214658: gap of 100 bp
* 214659 235423: contig of 20765 bp in length
* 235424 235523: gap of 100 bp
* 235524 283867: contig of 48344 bp in length
```

```

* 283968 283967: gap of 100 bp
* 283968 302319: contig of 18352 bp in length
* 302320 302419: gap of 100 bp
* 302420 309895: contig of 7476 bp in length
* 309896 309995: gap of 100 bp
* 309996 315745: contig of 5750 bp in length
* 315746 315845: gap of 100 bp
* 315846 333642: contig of 17797 bp in length
* 333643 333742: gap of 100 bp
* 333743 350641: contig of 16899 bp in length
* 350642 350741: gap of 100 bp
* 350742 365576: contig of 14835 bp in length
* 365577 365676: gap of 100 bp
* 365677 370201: contig of 4525 bp in length
* 370202 370301: gap of 100 bp
* 370302 377234: contig of 6933 bp in length
* 377235 377334: gap of 100 bp
* 377335 394656: contig of 17322 bp in length
* 394657 394756: gap of 100 bp
* 394757 410877: contig of 16121 bp in length.
* 394757 410877: Location/Qualifiers
FEATURES
source
1..410877
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-58J8"
/clone_id="Daniokey"
1..5913
/note="assembly fragment:00193
fragment chain:1"
misc_feature
6014..8737
/note="assembly fragment:00080
fragment chain:1"
misc_feature
8838..12348
/note="assembly fragment:00049
fragment chain:1"
misc_feature
12449..34866
/note="assembly fragment:01015
fragment chain:1"
misc_feature
34967..50012
/note="assembly fragment:00641
fragment chain:1"
misc_feature
50113..52494
/note="assembly fragment:00041
fragment chain:1"
misc_feature
52595..55884
/note="assembly fragment:00069
fragment chain:1"
misc_feature
55985..59774
/note="assembly fragment:00091
fragment chain:1"
misc_feature
59875..71704
/note="assembly fragment:00583
fragment chain:1"
misc_feature
71805..90091
/note="assembly fragment:00724
fragment chain:1"
misc_feature
90192..94302
/note="assembly fragment:00156
fragment chain:1"
misc_feature
94403..104620
/note="assembly fragment:00527
fragment chain:1"
misc_feature
104721..110745
/note="assembly fragment:00172
fragment chain:1"
misc_feature
110846..113873
/note="assembly fragment:00140
fragment chain:1"
misc_feature
113974..118974
/note="assembly fragment:00223
fragment chain:1"
misc_feature
119075..119540

```

```

/note="assembly fragment:01515
fragment chain:1"
misc_feature
139641..146697
/note="assembly fragment:00319
fragment chain:1"
misc_feature
146798..149204
/note="assembly fragment:00059
fragment chain:1"
misc_feature
149305..158380
/note="assembly fragment:00472
fragment chain:1"
misc_feature
158481..176802
/note="assembly fragment:00908
fragment chain:1"
misc_feature
176903..179349
/note="assembly fragment:00034
fragment chain:1"
misc_feature
179450..183727
/note="assembly fragment:00102
fragment chain:1"
misc_feature
183828..186517
/note="assembly fragment:00126
fragment chain:1"

Query Match 87.2%; Score 21.8; DB 14; Length 110000;
Best Local Similarity 92.0%; Pred. No. 3.2e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 77077 AAAAAAAAAACCTTCATCATTTAAAC 77101

RESULT 14
AL672217
LOCUS
DEFINITION
AL672217 117322 bp DNA linear VRT 16-APR-2005
zebrafish DNA sequence from clone BUSM1-72B14 Contains the gdf8
gene for growth differentiation factor 8, a novel gene similar to
PSM1 (postmeiotic segregation increased 1-like protein), two novel
genes, the 3' part of a novel gene, a novel glycoprotease, a novel
gene and two CpG islands, complete sequence.
ACCESSION
AL672217.6 GI:20338539
VERSION
HTG; CpG island; gdf8; glycoprotease; growth factor; PSM1.
KEYWORDS
SOURCE
Danio rerio (zebrafish)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 117322)
REFERENCE
Clark, G.
Direct Submission
Submitted (20-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonequests@sanger.ac.uk
On Apr 29, 2002 this sequence version replaced gi:20145727.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
http://www.sanger.ac.uk/Projects/C_elegans/wormpep BUSM1-72B14 is
from a Zebrafish PAC library
VECTOR: pCYPAC-6
This clone was isolated from BAC library (Genome Systems, St.
Louis) and provided by A. Donovon and L. Zon (Zon lab,
HMH/Children's Hospital of Boston, USA).
This sequence is the entire insert of clone BUSM1-72B14 This
sequence was finished as follows unless otherwise noted: all

```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_zerrio/fishmask.shtml

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zf1sh-help@sanger.ac.uk

FEATURES

Source

location/Qualifiers
1..117322
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="BUSM1-72814"
/clone_1lb="BUSM1"
complement(1)..50)
/note="Dr000298 repeat: matches 217. .265 of consensus"
70..199
/note="DNA15TA1_DR repeat: matches 133. .265 of consensus"
477..503
/note="2.5 copies 11 mer CTTATTACT 38% conserved"
641..651
/note="2.2 copies 5 mer TGAAG 22% conserved"
701..712
/note="3.0 copies 4 mer AAT 24% conserved"
763..775
/note="2.2 copies 6 mer AAATG 26% conserved"
929..938
/note="2.5 copies 4 mer ATGT 20% conserved"
1679..4527
/gene="gdf8"
join(1679..2133,2914..3281,4147..4527)
/gene="gdf8"
/product="SI:dZ72B14.1 (growth differentiation factor 8)"
/note="ZF1N record: ZDB-GENE-990415-165
match: CDNAS: Em:AJ318758 Em:AF019626"
/evidence=not_experimental
1721..1738
/note="2.2 copies 8 mer ATCAGAC 29% conserved"
join(1758..2133,2914..3281,4147..4527)
/gene="gdf8"
/note="ZF1N record: ZDB-GENE-990415-165
match: protein: Sw:O42222"
/evidence=not_experimental
/product="SI:dZ72B14.1 (growth differentiation factor 8)"
/protein_id="CAD43439.1"
/db_xref="GI:22204383"
/db_xref="GOA:O42222"
/db_xref="UniProt/Swiss-Prot:O42222"
/db_xref="ZF1N:ZDB-GENE-990415-165"
/translation="MHPTQVLSISVLIAAGVGYGITAQOPSTATTESSQCTE
FRGSKMLRLAIAKISQILSKRLKQAPNISRDVYKQILPKAPPIQQLIDQYDVGDS
KQGAVEEDHATITETITMTATBPDPVQVDRKRCFCFSFSPKIDQKRVTLWLVK
LRPAEATTVFLQISRLMPVVDGGRHRLRSIKIDVNGVTSWOSIDQKRVTLWLVK
ETNKGLEINADAGNDLAVTSTGEGLIPFMEVKSIEGPKIRVDSGLDCEENS
ESRCRYPLTVDFPDGMDVIAPRRYKANCSECEVDIMYQKPTHLLVKKASPRGT
AGCCPTPKMSINMLYNGKQIYIKIIPSMVYDRGCS"
complement(2608..2698)
repeat_region

/note="Dr000340 repeat: matches 138. .227 of consensus"
2766..2775
/note="2.5 copies 4 mer TATT 20% conserved"
3395..3405
/note="3.7 copies 3 mer AAT 22% conserved"
3610..3620
/note="2.8 copies 4 mer ACAT 22% conserved"
3920..3929
/note="2.5 copies 4 mer ATTT 20% conserved"
4019..4031
/note="2.2 copies 6 mer ACAATC 26% conserved"
4032..4041
/note="2.5 copies 4 mer ATTC 20% conserved"
4204..4214
/note="2.2 copies 5 mer GCACT 22% conserved"
4637..4647
/note="2.8 copies 4 mer CTCA 22% conserved"
4871..4885
/note="3.0 copies 5 mer AAAT 30% conserved"
5231..5241
/note="2.8 copies 4 mer ATGA 22% conserved"
5356..5367
/note="2.4 copies 5 mer ATTGT 24% conserved"
complement(join(5629..5972,10988..11028,12480..12667))
/gene="SI:dZ72B14.4"
complement(join(5629..5972,10988..11028,12480..12667))
/gene="SI:dZ72B14.4"
/product="SI:dZ72B14.4 (novel transcript)"
/note="match: ESTs: Bm:B1984224"
/evidence=not_experimental
5398..5947
/note="2.5 copies 4 mer TCGA 20% conserved"
6299..6310
/note="2.0 copies 6 mer AACAG 24% conserved"
7088..7104
/note="2.1 copies 8 mer TTCTCACT 34% conserved"
complement(7115..7374)
/note="dir591 repeat: matches 1. .250 of consensus"
complement(7376..7718)
/note="dir414 repeat: matches 1. .355 of consensus"
7758..7769
/note="12.0 copies 1 mer A 24% conserved"
7879..7892
/note="2.8 copies 5 mer TTAAT 28% conserved"
8004..8018
/note="2.1 copies 7 mer ATTATGA 30% conserved"
8269..8280
/note="2.0 copies 6 mer TCCAG 24% conserved"
8375..8387
/note="3.2 copies 4 mer CATT 26% conserved"
complement(8388..8612)
/note="HE1 DR1 repeat: matches 174. .384 of consensus"
8641..8673
/note="2.1 copies 16 mer AAGACTTTATTATC 50% conserved"
complement(8803..9059)
/note="HARBINGERNS_DR repeat: matches 595. .861 of
consensus"
complement(9065..9563)
/note="HARBINGERNS_DR repeat: matches 1. .489 of consensus"
9833..9858
/note="2.0 copies 13 mer TTGTTCTAAACTG 43% conserved"
9984..9994
/note="2.2 copies 5 mer TTAA 22% conserved"
10085..10107
/note="2.6 copies 9 mer AAGCAATA 37% conserved"
10653..10668
/note="3.2 copies 5 mer GTTTT 23% conserved"
complement(10738..10798)
/note="HA11 DR repeat: matches 577. .637 of consensus"
10808..10830
/note="5.8 copies 4 mer ATCC 46% conserved"
10828..10848
/note="5.2 copies 4 mer ATCT 24% conserved"
repeat_region

```

repeat_region      10965..10977
                    /note="3.2 copies 4 mer AAGC 26% conserved"
repeat_region      11034..11045
                    /note="6.0 copies 2 mer AC 24% conserved"
repeat_region      11241..11397
                    /note="Dr000396 repeat: matches 1..165 of consensus"
repeat_region      complement(11405..11512)
                    /note="Dr000358 repeat: matches 1..113 of consensus"
repeat_region      11533..11643
                    /note="Dr000038 repeat: matches 462..587 of consensus"
repeat_region      11682..11923
                    /note="TDR12 repeat: matches 2..242 of consensus"
repeat_region      12400..12410
                    /note="2.2 copies 5 mer AATTC 22% conserved"
repeat_region      12436..12450

Query Match      87.2%; Score 21.8; DB 5; Length 117322;
Best Local Similarity 92.0%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAAAAAAAACTTCATCATTTAAAC 25
        |||||
Db      56578 AAAAAAAAACTTCATCATTTGAAC 56602

RESULT 15
LOCUS      BS000225      134614 bp      DNA      linear      PRI 12-JUN-2004
DEFINITION Pan troglodytes chromosome 22 clone:PTB-062A13, map 22, complete
sequences.
ACCESSION      BS000225      BA000046
VERSION      BS000225.2      GI:38142448
KEYWORDS      HTG.
SOURCE      Pan troglodytes (chimpanzee)
            Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominoidea; Pan.
REFERENCE      1
            The international Chimpanzee Chromosome 22 Consortium.
            DNA sequence and comparative analysis of chimpanzee chromosome 22
            Nature 429, 382-388 (2004)
            2 (bases 1 to 134614)
            Park,H., Choi,S., Lee,Y., Park,K.H., Kim,H.K., Song,T., Joo,S.,
            Kim,K., Song,W., Chu,M., Jeong,J., Kim,N. and Kim,Y.
            Direct Submission
            Submitted (15-MAY-2003) Hong-Seog Park, Korea Research Institute of
            Bioscience and Biotechnology, Genome Research, 52 Oun-dong,
            Yuseong-gu, Daejeon, Chung Nam 305-333, Korea
            (E-mail:hspark@mail.kribb.re.kr, URL:http://chimp.kribb.re.kr,
            Tel:82-42-866-7181, Fax:82-42-860-4409)
            On Oct 31, 2003 this sequence version replaced gi:37537492.
            *The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
            *Chinese National Human Genome Center at Shanghai, Shanghai, China;
            *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
            of Molecular Biotechnology, Jena, Germany; *KRIIB Genome Research
            Center, Daejeon, Korea;
            *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
            *National Institute of Genetics, Mishima, Japan;
            *National Yang Ming University Genome Research Center, Taipei,
            Taiwan;
            *RIKEN Genomic Sciences Center, Yokohama, Japan.
            -----
            Center: KRIIB Genome Research Center
            Center code: KRIIB
            Web site: http://chimp.kribb.re.kr
            Contact: hspark@mail.kribb.re.kr
            -----
            Project Information
            Center project name:The Chimpanzee Chromosome 22 Sequencing Project
            Center clone name: PTB-062A13
            -----
            Summary Statistics
            Sequencing vector: pUC18,pUC13,PTZ19; 100% of reads Chemistry:
            Dye-terminator Big Dye and ET; 100% of reads Assembly program:

```

```

Phrap, version 0.990329
Consensus quality: 134,614 bases at least Q40
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.
-----
Source information:
The RPCI-43 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Cintc') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/mchimp43.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac).
VECTOR: pBACe3.6
The CHORI-251 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Cintc') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/chimpanzee251.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac).
VECTOR: pTABBAC2.1
The PTB1 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: PKS145
The PTF22 chimpanzee Fosmid library was prepared from DNA isolated
from cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: PKS143
-----
Sequence Quality Assessment:
This entry has been annotated with sequence
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in
10,000 bp.
-----
Neighboring clones: PTB-030G10(left) and PTB-057B23(right).
Location/Qualifiers
1..134614
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-062A13"
/clone_lib="PTB1 chimpanzee BAC"

FEATURES
source
Query Match      87.2%; Score 21.8; DB 8; Length 134614;
Best Local Similarity 92.0%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAAAAAAAACTTCATCATTTAAAC 25
        |||||

```

Db 114671 AAAAAAAAAATCTCACCATTAAAC 114695

RESULT 16
AC148528 141776 bp DNA linear PLN 06-NOV-2004

LOCUS AC148528
DEFINITION Medicago truncatula clone mth2-53h4, complete sequence.
AC148528
AC148528.15 GI:55468698
HTG.

VERSION
KEYWORDS
ORGANISM Medicago truncatula (barrel medic)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosidae; euroside I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE
AUTHORS 1 (bases 1 to 141776)
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
Medicago truncatula BAC Clone mth2-53h4
Unpublished

TITLE
JOURNAL 2 (bases 1 to 141776)
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (15-MAR-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE
AUTHORS 3 (bases 1 to 141776)
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (02-NOV-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

TITLE
JOURNAL 4 (bases 1 to 141776)
Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (06-NOV-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT
On Nov 6, 2004 this sequence version replaced gi:55168419.

Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
SOURCE
1. 141776
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-53h4"
/clone_1fb="Medicago truncatula BAC library H2"

ORIGIN
Query Match 87.2%; Score 21.8; DB 15; Length 141776;
Best Local Similarity 92.0%; Pred. No.3e+02; 2; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTAAC 25
DB 63123 AACAAATCTTCATCTTAAC 63147

RESULT 17
BS000165 142215 bp DNA linear PRI 12-JUN-2004
LOCUS BS000165
DEFINITION Pan troglodytes chromosome 22 clone:PTB-057B23, map 22, complete sequences.
BS000165 BA000046
ACCESSION BS000165.1 GI:37537432
VERSION

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HTG.
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryote; Metazoa; Chordata; Vertebrata; Eulestomni; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Pan.

1
The International Chimpanzee Chromosome 22 Consortium.
DNA sequence and comparative analysis of chimpanzee chromosome 22
Nature 429, 382-388 (2004)
2 (bases 1 to 142215)
Wang,S., Cai,Z., Wang,B., Zheng,H., Zhang,Y., Zhang,X., Zhu,G., Lu,G., Fu,G. and Chen,Z.
Direct Submission
Submitted (26-MAY-2003) Shengyue Wang, Chinese National Human Genome Center at Shanghai, Genomic Sequencing; No.250 BiBo Road, Zhang Jiang Hi-TECH Park, Shanghai 201203, CHINA
(E-mail: wangsy@chgc.sh.cn, URL: http://www.chgc.sh.cn, Tel:86-21-50801919, Fax:86-21-50801922)
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute of Molecular Biotechnology, Jena, Germany; *KRIIB Genome Research Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany; *National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei, Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.

Genome Center
Center: Chinese National Human Genome Center at Shanghai Center
code: CHGC
Web site: http://chgc.sh.cn
Contact: wangsy@chgc.sh.cn

Project Information
Center project name: The Chimpanzee Chromosome 22 Sequencing Project
Center clone name: PTB-057B23

Summary Statistics
Sequencing vector: pUC18, 100% of reads
Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly
program: Phrap; version 0.990329
Consensus quality: 142090 bases at least Q40
Consensus quality: 142207 bases at least Q30
Consensus quality: 142210 bases at least Q20
Quality coverage: 8.9x

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., paired quality >= 30) ;
an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at one plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.

Source information:
The PR1 chimpanzee BAC library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKS145
Sequence Quality Assessment:
This entry has been annotated with sequence
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Neighboring clones: PTB-062A13 (left) and RP43-094M15 (right).
Location/Qualifiers
1. 142215

```

/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-057B23"
/clone_11b="PTB1 chimpanzee BAC"

ORIGIN

Query Match      87.2%; Score 21.8; DB 8; Length 142215;
Best Local Similarity 92.0%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAACTGATCATTTAAAC 25
    |||
Db 19218 AAAAAAAAACTGATCATTTAAAC 19242

RESULT 18
BX323586      171172 bp      DNA      linear      VRT 14-JUL-2005
LOCUS      Zebrafish DNA sequence from clone CH211-303 in linkage group 9,
DEFINITION      complete sequence.
ACCESSION      BX323586
VERSION      BX323586.10 GI:70671328
KEYWORDS      HTG.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
      Cypriniformes; Cyprinidae; Danio.
      1 (bases 1 to 171172)
REFERENCE      Chapman, U.
AUTHORS      Direct Submission
      Submitted (13-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
      Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
      zfish-help@sanger.ac.uk
      http://www.sanger.ac.uk/Projects/D_rerio/fags.shtml#dateight
      On Jul 9, 2005 this sequence version replaced gi:68293323.
JOURNAL
TITLE      Genome Center
COMMENT      Center: Wellcome Trust Sanger Institute
      Center code: SC
      Web site: http://www.sanger.ac.uk
      Contact: zfish-help@sanger.ac.uk

-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep/
Zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.
Any regions longer than 1kb tagged as misc-feature 'unsure' are
part of a tandem repeat of more than 10kb in length where it has
not been possible to anchor the base differences between repeat
copies. The region has been built up based on the repeat element
to match the total size of repeat indicated by restriction digest,
but repeat copies may not be in the correct order and the usual
finishing criteria may not apply. CH211-303 is from a CHOR1-211 BAC
library
VECTOR: pTARBAC2.1.
Location/Qualifiers
1. 171172
/organism="Danio rerio"
/mol_type="genomic DNA"

```

```

/db_xref="taxon:7955"
/chromosome="9"
/clone="CH211-303"
/clone_11b="CHOR1-211"

ORIGIN

Query Match      87.2%; Score 21.8; DB 5; Length 171172;
Best Local Similarity 92.0%; Pred. No. 2.8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAACTGATCATTTAAAC 25
    |||
Db 157299 AAAAAAAAACTGATCATTTAAAC 157323

RESULT 19
AC164486/c      171984 bp      DNA      linear      HTG 29-JUL-2005
LOCUS      Danio rerio strain Tue clone CH211-31f19, WORKING DRAFT SEQUENCE, 8
DEFINITION      ordered pieces.
ACCESSION      AC164486
VERSION      AC164486.2 GI:71480250
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
      Cypriniformes; Cyprinidae; Danio.
      1 (bases 1 to 171984)
REFERENCE      Antonellis, A., Ayele, K., Bass, D., Benjamin, B., Bera, J.,
AUTHORS      Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Chu, G.,
      Coleman, H., Engle, J., Franks, S., Flinkenko, T., Geacole, M.,
      Greene, A., Guan, X., Gupta, J., Garrison, N., Haghighi, P., Han, J.,
      Hansen, N., Ho, S.-L., Hu, P., Hunter, G., Hulse, B., Idoi, J.R.,
      Kong, P., Laric, P., Larson, S., Lee-Lin, S.-Q., Legaspi, R.,
      Madden, M., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masetello, C.,
      Maskeri, B., McQuell, J., Mojidi, H.A., Mullikin, J.C., Park, M.,
      Portnoy, M.E., Prasad, A., Puri, O., Rantz, K., Reddix-Dugue, N.,
      Sante, A., Schandler, K., Schueler, M.G., Simon, C., Stantrop, S.,
      Tave, A., Thomas, J.W., Thomas, P.J., Tsipouri, V., Ung, L., Vogt, J.L.,
      Wetherby, K.D., Withers, T.R., Young, A. and Green, E.D.
      NISC Comparative Sequencing Initiative
      Unpublished
      2 (bases 1 to 171984)
REFERENCE      Green, E.D.
AUTHORS      Direct Submission
      Submitted (22-JUN-2005) NIH Intramural Sequencing Center, 5625
JOURNAL      Fishers Lane, Rockville, MD 20852, USA
      3 (bases 1 to 171984)
REFERENCE      Green, E.D.
AUTHORS      Direct Submission
      Submitted (29-JUL-2005) NIH Intramural Sequencing Center, 5625
JOURNAL      Fishers Lane, Rockville, MD 20852, USA
      On Jul 29, 2005 this sequence version replaced gi:68131651.
COMMENT      ----- Genome Center
      Center: NIH Intramural Sequencing Center
      Center code: NISC
      Web site: http://www.nisc.nih.gov
      Contact: nisc_zoo@nigrl.nih.gov
      ----- Project Information
      Center project name: em1
      Center clone name: 031f19

The sequence data in this record represents an 'enhanced'
version of a Phase 2 subdataset. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is generally based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence

```

contigs have been trimmed away, and each base is associated with a phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 169805 bases at least Q40
Consensus quality: 170526 bases at least Q30
Consensus quality: 170987 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 171284; sum-of-contigs
Quality coverage: 9.07x in Q20 bases; agarose-fp
Quality coverage: 10.48x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 6701: contig of 6701 bp in length
* 6702 6801: gap of unknown length
* 6802 43915: contig of 37114 bp in length
* 43916 44015: gap of unknown length
* 44016 56330: contig of 12315 bp in length
* 56331 56430: gap of unknown length
* 56431 87502: contig of 31072 bp in length
* 87503 87602: gap of unknown length
* 87603 115816: contig of 28214 bp in length
* 115817 115916: gap of unknown length
* 115917 118417: contig of 2500 bp in length
* 118418 167366: contig of 48850 bp in length
* 167367 167466: gap of unknown length
* 167467 171984: contig of 4518 bp in length.

FEATURES

source

1. 171984
/organism="Danio rerio"
/mol_type="genomic DNA"
/strain="Tue"
/db_xref="taxon:7955"
/clone="CH211-31p19"
/clone_lib="CH211"
/note="BAC resource: <http://bacpac.choi.org/>"
1. 6701
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
6702..6801
/estimated_length=unknown
6802..43915
/note="assembly_fragment"
43916..44015
/estimated_length=unknown
44016..56330
/note="assembly_fragment"
56331..56430
/estimated_length=unknown
56431..87502
/note="assembly_fragment"
87503..87602
/estimated_length=unknown
87603..115816
/note="assembly_fragment"
115817..115916
/estimated_length=unknown
115917..118416
/note="assembly_fragment"
118417..118516
/estimated_length=unknown
gap

misc_feature 118517..167366
/note="assembly_fragment"
167367..167466
/estimated_length=unknown
167467..171984
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN

Query Match 87.2%; Score 21.8; DB 14; Length 171984;
Best Local Similarity 92.0%; Pred. No. 2,8e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 40603 AAAAAAAAAACCTCATCTTAAC 25
AAAAAAAAACCTCATCTTGAAC 40579

RESULT 20

AC118168 191493 bp DNA linear HTG 15-NOV-2002
LOCUS
DEFINITION
Rattus norvegicus clone CH230-40608, WORKING DRAFT SEQUENCE.
AC118168
AC118168 GI:25009355
VERSION
HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 191493)
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Ayalabeche, V., Ayogagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blythe, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, J., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Dlye, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabist, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebreyegzi, E., Geier, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Huliy, S., Hume, J., Idelberg, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kover, C.,
Kowis, C., Kraft, C. L., Ledow, H., Levay, J., Lewis, L., Li, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshuwa, L., Louie, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Miolesavljevic, A., Miner, G., Ming, A., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokeneme, O., Okunolu, G., Olanrewaju, A., Pal, S., Parks, K.,
Paternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., L.,
Piazzi, M., Quirio, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smales, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sotelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,


```

TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 191493)
AUTHORS Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (14-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
JOURNAL 3 (bases 1 to 191493)
REFERENCE Rat Genome Sequencing Consortium.
TITLE Direct Submission
COMMENT Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:22856400.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold from the Atlas
assembly ('a' contig-scaffold'). Within each contig-scaffold,
individual sequence configs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
configs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only configs will be indicated in the feature
table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GYCK
Center clone name: CH230-406G8
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 180041 bases at least Q40
Consensus quality: 181059 bases at least Q30
Consensus quality: 181681 bases at least Q20
Estimated insert size: 183524; sum-of-configs estimation
Quality coverage: 7x in Q20 bases; sum-of-configs estimation

--- NOTE: Estimated insert size may differ from sequence length ---
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drift\_data.html)
* NOTE: This is a "working draft" sequence. It currently
* consists of 1 configs. Gaps between the configs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
+   1 191493: contig of 191493 bp in length.
Location/Qualifiers
+   1..191493
+   /organism="Rattus norvegicus"
+   /mol_type="Genomic DNA"
+   /db_xref="taxon:10116"
+   /clone="CH230-406G8"
+   ..1551
+   /note="wgs_end_extension
+   clone_end:T7"
misc_feature 7539..8410
FEATURES
SOURCE
misc_feature
misc_feature
misc_feature

```

```

ORIGIN
misc_feature
19058..191493
/Note="wgs_contig"

Query Match      87.2%; Score 21.8; DB 14; Length 191493;
Best Local Similarity 92.0%; Prid. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 AAAAAAAAAACCTTCATCATTTAAAC 25
|||||
Db      135686 AAGAAAAACCTTCATCATTTAAAC 135710

RESULT 21
CR792443/c
LOCUS      CR792443
DEFINITION zebrafish DNA sequence from clone DKEY-52K1 in linkage group 13,
complete sequence.
ACCESSION  CR792443
VERSION     CR792443.12 GI:72419832
KEYWORDS   HMG.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
AUTHORS     Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 217987)
            Wallis,J.
            Direct Submission
            Submitted (10-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk
            http://www.sanger.ac.uk/Projects/D_rerio/FAQ.shtml#dataeight
            On Aug 12, 2005 this sequence version replaced gi:72060589.
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -----

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Clone-derived zebrafish pUC subclones occasionally display
inconsistency over the length of mononucleotide A/T runs and
conserved TA repeats. Where this is found the longest good quality
representation will be submitted.
Any regions longer than 1kb tagged as misc-feature 'unsure' are
part of a tandem repeat of more than 10kb in length where it has
not been possible to anchor the base differences between repeat
copies. The region has been built up based on the repeat element
to match the total size of repeat indicated by restriction digest,
but repeat copies may not be in the correct order and the usual
finishing criteria may not apply.
DKEY-52K1 is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.
location/Qualifiers
1..217987
/organism="Danio rerio"
/mol_type="genomic DNA"

```

```

ORIGIN
Query Match      87.24; Score 21.8; DB 5; Length 217987;
Best Local Similarity 92.0%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

/db_xref="taxon:7955"
/chromosome="13"
/clone="DKEX-52X1"
/clone_lib="daniokey"

RESULT 22
AC098989/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-81J20, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
ACCESSION
AC098989
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 223743)
Muzny, D., Marie, Metcalf, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrook, S., Amth, A., Angiano, D.,
Anylabech, V., Ayogil, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Banahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, J., Foster, M., Foster, P.,
Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Girdy, M., Guerra, T., Guevara, W.,
Gutierrez, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howell, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovats, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, H., Louised, H., Lozano, R., Lu, X., Ma, J.,
Mageshwar, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martine, E.,
Mawhinley, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Mikolajewicz, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munida, M., Murphy, M., Nair, L.,
Narkewicz, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokwelen, O., Okwunu, G., Olariunagbon, A., Pal, S., Parks, K.,
Patel, S., Paul, H., Perez, A., Perez, L., Pflanz, C.,
Plopper, F., Polidexter, A., Popovic, D., Prilina, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shwartz, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced g1:24819417.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHUV
Center clone name: CH230-81J20
----- Summary Statistics
Assembly program: Atlas 3.0
Consensus quality: 198075 bases at least Q40
Consensus quality: 201307 bases at least Q30
Consensus quality: 203705 bases at least Q20
Estimated insert size: 209469; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be pre-empted.
*
* 1 219750: contig of 219750 bp in length
* 219751 219850: gap of unknown length
* 219851 221376: contig of 1526 bp in length
* 221377 221476: gap of unknown length
* 221477 223743: contig of 2267 bp in length.
Location/Qualifiers
1. 223743
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-81J20"
1. 1446
/notes="wgs end extension
clone_end:5pc"
3638..4289

```

```

misc_feature
misc_feature
misc_feature

```

```
misc_feature
          |
Query Match      87.2%; Score 21.8; DB 14; Length 223743;
Best Local Similarity 92.0%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORIGIN
gap
    gap
        /estimated_length=unknown
        22377..231476
        /estimated_length=unknown

Db      190927 AAGAAAAAACCCTCATCATTAAAC 190903
QY      1 AAAAAAAAAAACCCTCATCATTAAAC 25
         |||||
CR854945 231490 bp DNA linear VRT 19-MAR-2005
LOCUS zebrafish DNA sequence from clone DKEY-155H10 in linkage group 13,
DEFINITION complete sequence.
ACCESSION CR854945
VERSION CR854945
KEYWORDS CR854945.5 GI:61656927
HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
          1 (bases 1 to 231490)
          Henderson,C.
          Direct Submission
          Submitted (16-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
          On Mar 19, 2005 this sequence version replaced gi:60302501.
          ----- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfish-help@sanger.ac.uk
          -----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
validation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/WormPEP
Zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.
```

FEATURES	SOURCE
ORIGIN	<p>Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-155H10 is from a Zebrafish BAC library</p> <p>VECTOR: pindigBAC-5.</p> <p>Location/Qualifiers</p> <p>1. 231490</p> <p>/organism="Danio rerio"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:7955"</p> <p>/clone="DKEY-155H10"</p> <p>/clone_1lb="DanioKey"</p>
Query Match	87.2%; Score 21.8; DB 5; Length 231490;
Beet Local Similarity	92.0%; Pred. No. 2.5e+02;
Matches	23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy	<p>1 AAAAAAAAAACCTCATTTTAAAC 25</p> <p> </p> <p>AAAAAAAAAACCTCATTCGTTTAAAC 229872</p>
Db	229896 AAAAAAAAAACCTCATTCGTTTAAAC 229872
RESULT 24	
AC129699/c	
LOCUS	AC129699 300486 bp DNA linear HTG 09-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-11122. *** SEQUENCING IN PROGRESS
ACCESSION	AC129699
VERSION	AC129699.4 GI:24818133
KEYWORDS	HTG; HTGS PHASE1; HTGS DNAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurgnathi; Muridae; Muridae; Murinae; Rattus.
AUTHORS	1 (bases 1 to 300486)
	Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biewald,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinj,H, DiVya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Gante,R, Garcia,R, Garner,T, Garza,M, Georgescu,E, Geer,K, Gall,R, Grady,M, Guerra,M, Guevara,W, Gunaratne,P, Haaland,M, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,B, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huylk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowic,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenzowa,L, Louisedge,H, Lozado,R,J, Lu,X, Ma,J, Maheswari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,B, Mashinsky,S, McLeod,M,P, McNeill,T,Z, Meenan,E, Milosavljevic,A, Miner,G, Mintz,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwaokwelenh,O, Okwuonu,G, Olarinpoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfamkoch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L,-L,

VERSION BX548026.10 GI:45434023
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 147902)
 AUTHORS Barker, G.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAR-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
 COMMENT On Mar 14, 2004 this sequence version replaced gi:44644515.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. Clone-derived zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
 Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml. CH211-193D9 is from a CHORI-211 BAC library
 VECTOR: pTARBAC2.1.
 FEATURES
 source Location/Qualifiers
 1..147902
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-193D9"
 /clone_id="CHORI-211"
 ORIGIN
 Query Match 85.6%; Score 21.4; DB 5; Length 147902;
 Best Local Similarity 95.7%; Pred. No. 4e+02;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTTCATCATTTTA 23
 Db 34708 AAAAAAAAACTTCATCATTTTA 34730
 RESULT 26
 LOCUS AL607108 202842 bp DNA linear ROD 10-APR-2002
 DEFINITION Mouse DNA sequence from clone RP23-42F6 on chromosome 11, complete sequence.

ACCESSION AL607108
 VERSION AL607108.17 GI:20145308
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Brown, J.
 TITLE Direct Submission
 JOURNAL Submitted (10-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquer@sanger.ac.uk
 COMMENT On Apr 12, 2002 this sequence version replaced gi:19849646.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. RP23-42F6 is from the RP23 Mouse PAC library constructed by the group of Pletier de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6.
 FEATURES
 source Location/Qualifiers
 1..202842
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-42F6"
 /clone_id="RP23-42F6"
 ORIGIN
 Query Match 84.0%; Score 21; DB 9; Length 202842;
 Best Local Similarity 100.0%; Pred. No. 5e+02;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTTCATCATTT 21
 Db 87969 AAAAAAAAACTTCATCATTT 87989
 RESULT 27
 LOCUS AX345821 5203 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 892 from Patent WO0200928.
 AX345821
 VERSION AX345821.1 GI:18493707
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 other sequences; artificial sequences.
 REFERENCE 1
 AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
 TITLE Diagnosis of diseases associated with the immune system
 JOURNAL Patent: WO 0200928-A 892 03-JAN-2002;
 EpiGenomics AG (DE)
 FEATURES
 source Location/Qualifiers
 1..5203

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 5203;
Best Local Similarity 91.7%; Pred. No. 2e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTTCATCTTAAC 25
Db 2966 AAAAAATACCTCATATTAAAC 2943

RESULT 28
CEM05B5/C 29482 bp DNA linear INV 09-AUG-2005
DEFINITION Caenorhabditis elegans Cosmid W05B5, complete sequence.
ACCESSION Z82071
VERSION Z82071.1 GI:3217782
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;
Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 29482)
REFERENCE 1
AUTHORS C. elegans Sequencing Consortium
CONSRM Genome sequence of the nematode C. elegans: a platform for
TITLE investigating biology
JOURNAL Science 282 (5396), 2012-2018 (1998)
PUBMED 9851916
2 (bases 1 to 29482)
AUTHORS McLay K.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
worm@sanger.ac.uk
COMMENT On Jun 13, 1998 this sequence version replaced gi:1666032.
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring subclones.
For a graphical representation of this sequence and its analysis
see: [http://www.wormbase.org/perl/ace/elegans/seq/sequence?](http://www.wormbase.org/perl/ace/elegans/seq/sequence?name=W05B5;class=Sequence)
[name=W05B5;class=Sequence](http://www.wormbase.org/perl/ace/elegans/seq/sequence?name=W05B5;class=Sequence)
This sequence is the entire insert of clone W05B5. The true right
end of clone ZK256 is at 14999 in this sequence. The start of this
sequence (1..105) overlaps with the end of sequence Z82088.
The end of this sequence (29376..29482) overlaps with the start of
sequence AL110478.

FEATURES
SOURCE Location/Qualifiers
1..29482
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="W05B5"
join(113084..13199,13367..13498,14205..14360,15105..15236,
15415..15508)
/locus_tag="W05B5.1"
join(13084..13199,13367..13498,14205..14360,15105..15236,
15415..15508)

gene 15415..15508)
/locus_tag="W05B5.1"
/standard_name="W05B5.1"
/note="contains similarity to Saccharomyces cerevisiae
Homolog of human WASP, proline-rich protein; SGD:YOR181W"

CDS /codon_start=1
/product="Hypothetical protein W05B5.1"
/protein_id="CAB04917.1"
/db_xref="GI:3880497"
/db_xref="UniProt/TREMBL:Q9XUK9"
/translation="MSLWGFDVTEVLSAFTIGENPMAYFQVFEVCWCFIPDAPV
SRVVTSDVTYEPAPSPIPPAPQPEYRRAVAPPPPSKMTAENPFRBEIYHV
DPVEQYLHKPPPPSRGSAQNTPTQOHTQATPTPPNHSPLTQNSLEQOLQ
NEKNEGHNSSEPLASHNREYEDLPKAYVELLHVKKKCGCSVO"
join(20852..20994,21666..21753,23781..23927,23980..24094,
24139..24233,24435..24589,24845..24955,25420..25654)
/locus_tag="W05B5.2"
join(20852..20994,21666..21753,23781..23927,23980..24094,
24139..24233,24435..24589,24845..24955,25420..25654)
/locus_tag="W05B5.2"
/standard_name="W05B5.2"
/note="contains similarity to Pfam domain PF00001 (7
transmembrane receptor (rhodopsin family))"
/codon_start=1
/product="Hypothetical protein W05B5.2"
/protein_id="CAB04918.2"
/db_xref="GI:70720750"
/db_xref="GOA:Q9XUK8"
/db_xref="UniProt/TREMBL:Q9XUK8"
/translation="MLDYNVNTSGSSEEDVDTSEITGLQVTHREYEDITAAELMPVA
VEQMPYITLASHMNVIGYIGTLVYVYVATKSRNALNLVNLAIADLLILFCLPL
TVNDVYKTFWFSVAVFCKSVFNNTSVVYSINSLVFTERRALITYPKSPVRR
SVNGIWFILAMPSPSPVTLHAGAPFVAPNPTTGTGCKSMBEFPQKQLOLT
IFSEVPLVLVISILCLMVRTLHFSANVLTVARQISIRKAVRMCAVFLFSMVL
PVLNVALNLYDLSTDSVNTTAVRKLPRVVSYSCLNPILYSFLSGRFECISY
ILKNSIKPTDIYITLKQTNKSRK"
join(complement(AL110478.1:2838..3117),
complement(AL110478.1:520..641),
complement(AL110478.1:108..136),complement(29367..29482),
complement(28407..28570),complement(28177..28326),
complement(27442..27813),complement(26542..26853))
/gene="nhr-85"
/locus_tag="W05B5.3a"
join(complement(AL110478.1:2838..3117),
complement(AL110478.1:520..641),
complement(AL110478.1:108..136),complement(29367..29482),
complement(28407..28570),complement(28177..28326),
complement(27442..27813),complement(26542..26853))
/gene="nhr-85"
/locus_tag="W05B5.3a"
/standard_name="W05B5.3a"
/note="C. elegans NHR-85 protein; contains similarity to
Pfam domains PF00104 (ligand-binding domain of nuclear
hormone receptors), PF00105 (zinc finger, C4 type (two
domains))"
/codon_start=1
/product="Hypothetical protein W05B5.3a"
/protein_id="CAB04919.2"
/db_xref="GI:11060535"
/db_xref="GOA:Q9XUK7"
/db_xref="HSSP:1A6Y"
/db_xref="InterPro:IPR000324"
/db_xref="InterPro:IPR000516"
/db_xref="InterPro:IPR001628"
/db_xref="InterPro:IPR001723"
/db_xref="InterPro:IPR008946"
/db_xref="UniProt/Swiss-Prot:Q9XUK7"
/translation="WDLSTSGSLPCHPLPLSLVLTNQPPIDTSRSPAPSSGSLSP
SPSAPFVVPKSLMNPQAMPDTFYMSALISLVLTNDQVSHGLKESKSLAKB
NASTSAGTILICVCSDAVSGDFHVPFQCECKGFFRSIOKLTYPKTAEDCLLR
NNRNRCCKRLKCKCAVMSGDVAFVPRVREKAFMFEQKTNVQSGQDLIOYE
NLTEVMKINQAGTLOATLEKCTGPTLTORCPITTSFVIPLKAAIDFANISAFS
ITOTQRVHLIQNSVFVDMLASASASISQHPFPGGLTYDDSSANPIIPQALQISARI

gene	ROLPOTVLTATAVCOADLLPESQPMILAEKMCVGLGGLGIGSLATPRLADY RTLRQMSDRKQMSQISQHSQNLILPVAAPVLLPAPLSPAPASVSSSVK SEFLERKPSLISLERPRRISSGAQEPLNLSLPHVRHQVRDVSDEQLEMKRSPV PTLISE"			
	complement {join(26542..26853,27442..27813,28177..28326, 28407..28570,29367..29379)} /gene="hnr-85" /locus_tag="W05B5.3b" complement {join(26542..26853,27442..27813,28177..28326, 28407..28570,29367..29379)} /gene="hnr-85" /locus_tag="W05B5.3b" /standard_name="W05B5.3b" /codon_start=1 /product="Hypothetical protein W05B5.3b" /protein_id="CAH04734.1" /db_xref="GI:50507784" /db_xref="GOA:Q6BER3" /db_xref="InterPro:IPR000536" /db_xref="UniProt/TREMBL:Q6BER3" /translation="MSRDVAFGVPRKREKRMFEEMQKTNVQSGORDQIAQYENLTE VMKINQAFGLQATLEKCTGPIYTRCPITSNIVIPLKAIIDPANSIPAFSLITOT QRVHLQNSVDFVWLASASASTSQHPGGGLTYDOSANPILPOAISARIRQLP POTVLTATAVCOADLLPESQPMILAEKMCVGLGGLGIGSLATPRLADYRTLR QMSDRKQMSQISQHSQNLILPVAAPVLLPAPLSPAPASVSSSVKSEFLERKPSLIS LERPRRISSGAQEPLNLSLPHVRHQVRDVSDEQLEMKRSPVPTLISE"			
CDS				
ORIGIN				
Query Match	83.2%;	Score 20.8;	DB 2;	Length 29482;
Best Local Similarity	91.7%;	Pred. No. 1.1e+03;		
Matches 22;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1 AAAAAAAAAACCTCATCTTAA 24 24997 AAAAAAAAAACTTCACATTAA 24974			
Db	24997 AAAAAAAAAACTTCACATTAA 24974			
RESULT 29				
U00046/c	U00046 35396 bp DNA linear INV 30-JUL-2005			
LOCUS	Caenorhabditis elegans cosmid R13f6, complete sequence.			
DEFINITION	U00046			
ACCESSION	U00046.1 GI:470358			
VERSION	HTG.			
KEYWORDS	Caenorhabditis elegans			
SOURCE	Caenorhabditis elegans			
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Pelodermidae; Caenorhabditis.			
REFERENCE	1 (bases 1 to 35396)			
AUTHORS	C. elegans Sequencing Consortium			
CONSTRM	Genome sequence of the nematode C. elegans: a platform for			
TITLE	investigating biology			
JOURNAL	Science 282 (5396), 2012-2018 (1998)			
PUBMED	9651916			
REFERENCE	2 (bases 1 to 35396)			
AUTHORS	Miller,N.			
CONSTRM	The sequence of C. elegans cosmid R13f6			
TITLE	Unpublished (2001)			
JOURNAL	3 (bases 1 to 35396)			
REFERENCE	Waterston,R.			
AUTHORS	Direct Submission			
CONSTRM	Submitted (08-APR-1994) Department of Genetics, Washington			
TITLE	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
JOURNAL	Louis, MO 63110, USA			
REFERENCE	4 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (28-MAY-1996) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	5 (bases 1 to 35396)			
AUTHORS	Louis, MO 63110, USA			
CONSTRM	Submitted (30-JUL-2005) Department of Genetics, Washington			
TITLE	Submitted (30-JUL-2005) Department of Genetics, Washington			
JOURNAL	Submitted (30-JUL-2005) Department of Genetics, Washington			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (09-AUG-2001) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	6 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (21-SEP-2001) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	7 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (23-MAY-2002) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	8 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (28-AUG-2002) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	9 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (13-NOV-2002) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	10 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (26-DEC-2002) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	11 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (10-FEB-2003) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	12 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (07-APR-2003) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	13 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (21-SEP-2004) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	14 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (15-APR-2005) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	15 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (19-MAY-2005) Department of Genetics, Washington			
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.			
REFERENCE	16 (bases 1 to 35396)			
AUTHORS	Waterston,R.			
CONSTRM	Direct Submission			
TITLE	Submitted (30-JUL-2005) Department of Genetics, Washington			
JOURNAL	Submitted (30-JUL-2005) Department of Genetics, Washington			

COMMENT

University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: submissions@watson.wustl.edu and jee@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this clone sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=R13f6;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is F28F5, 200 bp overlap; the 3' clone is K04C2, 1000 bp overlap.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yujii Kohara (http://www.ddb.jhg.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://wormfdb.dcfcl.harvard.edu/>), similarity to other proteins from Blast analysis (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SF (Lowe, T.M. and Bddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

CDS

gene

```
1..35396
  /organism="Caenorhabditis elegans"
  /mol_type="genomic DNA"
  /strain="Bristol N2"
  /db_xref="taxon:6239"
  /chromosome="III"
  /clone="R13f6"
  3387..11863
    /gene="tag-152"
    /locus_tag="R13f6_4"
    join(3387..3541,3605..3786,3878..4049,4101..5371,
    5424..6668,6716..7231,7284..8383,8447..8803,9050..10970,
    11019..11670,11752..11863)
    /gene="tag-152"
    /locus_tag="R13f6_4"
    /standard_name="R13f6_4a"
    /note="contains similarity to Pfam domains PF05593 (RMS Repeat), PF00008 (EGF-like domain), PF01436 (NH2 repeat);
    coded for by the following C. elegans cDNAs: yk33d1.5,
    yk194c6.5, yk209c11.5, yk240f12.5, yk357a10.5, yk529b.5,
    yk589d10.5, yk625d3.5, yk627g3.5, yk890e07.3, yk890e07.5,
    yk1010e03.5, yk1281a10.5, yk1467b01.3, yk1467b01.5,
    CBERM90F, CBERM39F, yk5b4.3, yk194c6.3, yk209c11.3,
    yk357a10.3, yk625d3.3, yk627g3.3, yk1281a10.3"
    /product="Temporarily assigned gene name protein 152,"
```

Query Match 83.2%; Score 20.8; DB 2; Length 35396;
Best Local Similarity 91.7%; Pred. No. 1.le+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 21169 AAAAAATAACCTTCATCATTTAA 21146

RESULT 30
AX695389/c 37487 bp DNA linear PAT 31-MAR-2003
LOCUS AX695389 Sequence 1016 from Patent WO03008583.
DEFINITION AX695389
ACCESSION AX695389
VERSION AX695389.1 GI:29418539
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Morris, D.W. and Engelhard, E.K.
TITLE Novel compositions and methods for cancer
JOURNAL Patent: WO 03008583-A 1016 30-JAN-2003;
Sages Discovery (US)
FEATURES
source
1..37487
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 37487;
Best Local Similarity 91.7%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 36890 AAAAAACATACCTTCATCATTTAA 36867

RESULT 31
AC016834/c 66440 bp DNA linear HTG 13-JUL-2000
LOCUS AC016834 Homo sapiens clone R11-1U13, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC016834
ACCESSION AC016834
VERSION AC016834.2 GI:9119767
KEYWORDS HTG; HTGS_PHASED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 66440)
Britten, B., Linton, L., Nussbaum, C. and Lander, E.
Homo sapiens, clone RP11-1J13
Unpublished

2 (bases 1 to 66440)
Britten, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barne, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,
Leloczky, J., Lien, C., Locke, K., Macdonald, P., Margulis, N.,
McGowan, P., McGurk, A., McKernan, K., McDonald, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Strange-Thomann, N., Stojanovic, N., Sudramanian, A., Talamas, J.,
Teafey, S., Tittell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6539375.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2498

Center clone name: 1_U_13

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
813 812: contig of 812 bp in length
913 912: gap of 100 bp
1728 1727: contig of 815 bp in length
1828 1827: gap of 100 bp
2636 2635: contig of 808 bp in length
2736 2735: gap of 100 bp
3544 3544: contig of 809 bp in length
3645 3644: gap of 100 bp
4497 4496: contig of 852 bp in length
4597 4596: gap of 100 bp
5409 5408: contig of 812 bp in length
5509 5508: gap of 100 bp
6314 6313: contig of 805 bp in length
7191 7190: contig of 777 bp in length
7291 7290: gap of 100 bp
8101 8100: contig of 810 bp in length
8201 8200: gap of 100 bp
8977 8976: gap of 100 bp
9077 9076: contig of 776 bp in length
9887 9887: contig of 811 bp in length
9988 9987: gap of 100 bp
10776 10776: contig of 789 bp in length
10777 10876: gap of 100 bp

10877 11779: contig of 903 bp in length
11780 11879: gap of 100 bp
11880 12694: contig of 815 bp in length
12695 12794: gap of 100 bp
12795 13590: contig of 796 bp in length
13591 13690: gap of 100 bp
13691 14501: contig of 811 bp in length
14502 14601: gap of 100 bp
14602 15387: contig of 786 bp in length
15388 15487: gap of 100 bp
15488 16314: contig of 827 bp in length
16315 16414: gap of 100 bp
16415 17222: contig of 808 bp in length
17223 17322: gap of 100 bp
17323 18141: contig of 819 bp in length
18142 18241: gap of 100 bp
18242 19047: contig of 806 bp in length
19048 19147: gap of 100 bp
19148 19976: contig of 829 bp in length
19977 20076: gap of 100 bp
20077 20902: contig of 826 bp in length
20903 21002: gap of 100 bp
21003 21821: contig of 819 bp in length
21822 21921: gap of 100 bp
21922 22717: contig of 796 bp in length
22718 22817: gap of 100 bp
22818 23640: contig of 823 bp in length
23641 23740: gap of 100 bp
23741 24573: contig of 833 bp in length
24574 24673: gap of 100 bp
24674 25458: contig of 785 bp in length
25459 25558: gap of 100 bp
25559 26385: gap of 100 bp
26385 26484: contig of 826 bp in length
26485 27293: contig of 809 bp in length
27294 27393: gap of 100 bp
27394 28207: contig of 814 bp in length
28208 28307: gap of 100 bp
28308 29114: contig of 807 bp in length
29115 29214: gap of 100 bp
29215 30036: contig of 822 bp in length
30037 30137: gap of 100 bp
30137 30939: contig of 803 bp in length
30939 31039: gap of 100 bp
31040 31864: contig of 825 bp in length
31865 31964: gap of 100 bp
31965 32763: contig of 799 bp in length
32764 32863: gap of 100 bp
32864 33651: contig of 788 bp in length
33652 33751: gap of 100 bp
33752 34538: contig of 787 bp in length
34539 34638: gap of 100 bp
34639 35443: contig of 805 bp in length
35444 35543: gap of 100 bp
35544 36353: contig of 810 bp in length
36354 36453: gap of 100 bp
36454 37260: contig of 807 bp in length
37261 37360: gap of 100 bp
37361 38195: contig of 835 bp in length
38196 38295: gap of 100 bp
38296 39124: contig of 829 bp in length
39125 39224: gap of 100 bp
39225 40050: contig of 826 bp in length
40051 40150: gap of 100 bp
40151 40977: contig of 827 bp in length
40978 41077: gap of 100 bp
41078 41869: contig of 792 bp in length
41870 41969: gap of 100 bp
41970 42780: contig of 811 bp in length
42781 42880: gap of 100 bp
42881 43670: contig of 790 bp in length
43671 43770: gap of 100 bp
43771 44595: contig of 825 bp in length

```

* 44596 44695: gap of 100 bp
* 44696 45513: contig of 818 bp in length
* 45514 45613: gap of 100 bp
* 45614 46514: contig of 844 bp in length
* 46514 46557: gap of 100 bp
* 46558 47368: contig of 811 bp in length
* 47369 47469: gap of 100 bp
* 47470 48274: contig of 806 bp in length
* 48275 48374: gap of 100 bp
* 48375 49185: contig of 811 bp in length
* 49186 49285: gap of 100 bp
* 49286 50099: contig of 814 bp in length
* 50100 50199: gap of 100 bp
* 50200 51005: contig of 806 bp in length
* 51006 51907: contig of 802 bp in length
* 51908 52007: gap of 100 bp
* 52008 52831: contig of 824 bp in length
* 52832 52931: gap of 100 bp
* 52932 53751: contig of 819 bp in length
* 53752 54665: gap of 100 bp
* 54666 54765: gap of 100 bp
* 54766 55567: contig of 801 bp in length
* 55568 56477: gap of 100 bp
* 56478 56577: gap of 100 bp
* 56578 57398: contig of 821 bp in length
* 57399 57498: gap of 100 bp
* 57499 58281: contig of 783 bp in length
* 58282 58381: gap of 100 bp
* 58382 59187: contig of 806 bp in length
* 59188 59287: gap of 100 bp
* 59288 60085: contig of 798 bp in length
* 60086 60185: gap of 100 bp
* 60186 60996: contig of 811 bp in length
* 60997 61096: gap of 100 bp
* 61097 61925: contig of 829 bp in length
* 61926 62025: gap of 100 bp
* 62026 62845: contig of 820 bp in length
* 62846 62945: gap of 100 bp
* 62946 63766: contig of 821 bp in length
* 63767 63866: gap of 100 bp
* 63867 64655: contig of 789 bp in length
* 64656 64755: gap of 100 bp
* 64756 65548: contig of 793 bp in length
* 65549 65648: gap of 100 bp

```

Query Match 83.2%; Score 20.8; DB 14; Length 66440;
 Best Local Similarity 91.7%; Pred. No. 8.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
 Db 33463 AAAAAAAAACTTCATTTAA 33440

RESULT 32
 AL162491/c
 LOCUS Human DNA sequence from clone RP11-103D20 on chromosome 13 Contains
 DEFINITION a Cpg Island, complete sequence.
 ACCESSION AL162491
 VERSION AL162491.10 GI:1043407
 KEYWORDS HTG; CPG Island.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 1 (bases 1 to 79374)
 Holte,K.
 Direct Submission

JOURNAL
 Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
 Clone requests: clonerequests@sanger.ac.uk
 On Oct 1, 2000 this sequence version replaced gi:10045368.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr13
 RP11-103D20 is from the library RP11-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBAC3.6

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one subclone; and the assembly was confirmed by restriction digest,
 except on the rare occasion of the clone being a YAC.

FEATURES

source

1. 79374
 Location/Qualifiers

/organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-103D20"
 /clone_1id="RP11-11.1"
 1
 /note="Clone_left_end: RP11-103D20"
 79275
 /note="Clone_left_end: RP11-255L12"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 79374;
 Best Local Similarity 91.7%; Pred. No. 8e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
 Db 25694 AAAAAAAAACTTCATTTAA 25671

RESULT 33
 AC164748
 LOCUS Bos taurus clone CH240-155M12, *** SEQUENCING IN PROGRESS ***
 DEFINITION unordered pieces.
 ACCESSION AC164748
 VERSION AC164748.1 GI:68265676
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 100161)
 Murzyn,D., Adams,C., Agbai II,O., Allen,C., Albrooks,S., Archer,P.,
 Arredondo,H., Barendse,D., Bangura,J., Beltran,B., Beltran,R.,
 Beraducci,A., Biswal,K., Blyth,P., Boham,H., Bunay,C., Burch,P.,
 Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
 Chacko,J., Chahour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
 Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,

Curry, S., Dai, W., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Donlin, J., McCauley, S., Dugan-Rocha, S., Dunn, A., Durbin, K., Dzunda, D., Egan, A., Escoto, M., Espinosa, V., Eugene, C., Fa, M., Fernandez, S., Fernando, P., Flagg, N., Forbes, L., Foster, P., Fowler, G., Fu, O., Fuh, E., Garcia, A., Garcia, R., Garner, T., Gaskin, C., Gensch, S., Ghouse, S., Gill, R., Gonzalez, D., Gonzalez-Garay, M., Guevara, W., Holder, M., Haaland, W., Haebler, K., Hall, B., Hamid, H., Hamilton, K., Harbes, B., Harris, R., Havlak, P., Haves, A., Hawkins, E., Hayes, S., Hemphill, L., Hernandez, J., Hites, S., Hitchens, M., Hodgson, A., Hognes, M., Hollins, B., Howell, L. T., Huylk, S., Hume, J., Imo, K., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Kafatis, K., Kelly, S., Key, T., Khan, Z., King, L., Koyar, C., Kowis, A., Kowis, C., Lala, F., Lee, S., Lee, K., Lee, S., Legall, F. I., Lemon, S., Lewis, L., Li, B., Li, Y., Li, Z., Linneil, M., Liu, W., Liu, Y. S., Liu, Y., Liyanage, D., London, P., Lopez, J., Lorenshewa, L., Lozardo, R., Luk, T., Madu, R., Maheshwari, M., Mahoney, C., Malloy, K., Mansouri, D., Martinez, E., McLelland, H., McPherson, J., Mercadeo, C., Metzger, M., Milosavljevic, A., Minja, E., Morgan, M., Morris, S., Muidasa, M., Murray, D., Nazareth, L., Ngo, D., Nguyen, N., Norwig-Eastaugh, E., Nott, A., Nwaokelemeh, O., Oregon, M., Ochi-Okorie, C., Odeh, E., Okwuonu, G., Okwuonu, K., Parker, D., Pasternak, S., Patel, B., Patel, V., Paul, H., Perez, A., Perez, L., Petrosino, J., Pham, T., Primus, E., Pu, L., Puazo, M., Qin, X., Quinn, A., Quiroz, J., Rabata, D., Rachlin, E., Reigh, R., Ren, Y., Reuter, M., Richards, S., Rivers, C., Rodriguez, F., Rojas, A., Ruiz, S. J., Sana, M., Sanders, W., Santibanez, J., Santos, R., Savary, G., Scherer, S., Shen, H., Shen, Y., Sisson, I., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Svatek, A., Taylor, E., Taylor, T., Thomas, N., Thorn, R., Thornton, R., Trejos, Z., Umami, K., Vargo, C., Verdusco, D., Villaseana, D., Virk, D., Volkov, A., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, R., Wei, X., Wheeler, D., Williams, G., Williams, R., Worley, K., Wright, R., Wu, J., Yakub, S., Yan, K., Yuan, Y., Yu, F., Zhang, J., Zhang, L., Zhang, Z., Zhou, J., Weinstein, G. and Gibbs, R.

Unpublished
Direct Submission
2 (bases 1 to 100161)
Worley, K.C.
Direct Submission
Submitted (28-JUN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help.tmc.edu

----- Project Information -----
Center project name: F0DF
Center clone name: CH240-155M12
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 103355 bases at least Q40
Consensus quality: 108736 bases at least Q30
Consensus quality: 111968 bases at least Q20
Estimated insert size: 123072; sum-of-coverage estimation
Quality coverage: 2x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5131: contig of 5131 bp in length
* 5132 5231: gap of unknown length

FEATURES	SOURCE
5232	8313: contig of 3082 bp in length
8314	8413: gap of unknown length
8414	11909: contig of 3496 bp in length
11910	12009: gap of unknown length
12010	17131: contig of 5122 bp in length
17132	17231: gap of unknown length
17232	21787: contig of 4556 bp in length
21788	21887: gap of unknown length
21888	21995: contig of 2108 bp in length
23996	24095: gap of unknown length
24096	26495: contig of 2400 bp in length
26496	26595: gap of unknown length
26596	28813: contig of 2218 bp in length
28814	28913: gap of unknown length
28914	32745: contig of 3832 bp in length
32746	32845: gap of unknown length
32846	35067: contig of 2222 bp in length
35068	35167: gap of unknown length
35168	38311: contig of 3144 bp in length
38312	38411: gap of unknown length
38412	41558: contig of 3147 bp in length
41559	41658: gap of unknown length
41659	43804: contig of 2146 bp in length
43805	43904: gap of unknown length
43905	46068: contig of 2164 bp in length
46069	46168: gap of unknown length
46169	48555: contig of 2387 bp in length
48556	48655: gap of unknown length
51261	51261: contig of 2606 bp in length
51362	51361: gap of unknown length
56689	56689: contig of 5328 bp in length
56790	56789: gap of unknown length
56790	56966: contig of 2907 bp in length
56967	59796: gap of unknown length
59797	62036: contig of 2240 bp in length
62037	62136: gap of unknown length
62137	65659: contig of 3523 bp in length
65660	65759: gap of unknown length
65760	68284: contig of 2525 bp in length
68285	68384: gap of unknown length
68385	70705: contig of 2321 bp in length
70706	70805: gap of unknown length
70806	73031: contig of 2226 bp in length
73032	73131: gap of unknown length
73132	78880: contig of 5749 bp in length
78881	78980: gap of unknown length
82073	82072: contig of 3092 bp in length
82173	82172: gap of unknown length
84604	84603: contig of 2431 bp in length
84704	84703: gap of unknown length
84704	87435: contig of 2732 bp in length
87436	87535: gap of unknown length
87536	93034: contig of 5499 bp in length
93035	93134: gap of unknown length
93135	97117: contig of 3983 bp in length
97118	97217: gap of unknown length
97218	100161: contig of 2944 bp in length.
100161	Location/Qualifiers
1. 100161	/organism="Bos taurus"
	/mol_type="genomic DNA"
	/db_xref="taxon:9913"
	/clone="CH240-155M12"
5132. 5231	
	/estimated_length=unknown
8314. 8413	
	/estimated_length=unknown
11910. 12009	
	/estimated_length=unknown
17132. 17231	
	/estimated_length=unknown
21788. 21887	
	/estimated_length=unknown

gap 23996. .24095
/estimated_length=unknown
gap 26496. .26595
/estimated_length=unknown
gap 28814. .28913
/estimated_length=unknown
gap 32746. .32845
/estimated_length=unknown
gap 35068. .35167
/estimated_length=unknown
gap 38312. .38411
/estimated_length=unknown
gap 41559. .41658
/estimated_length=unknown
gap 43805. .43904
/estimated_length=unknown
gap 46069. .46168
/estimated_length=unknown
gap 48556. .48655
/estimated_length=unknown
gap 51262. .51361
/estimated_length=unknown
gap 56690. .56789
/estimated_length=unknown
gap 59697. .59796
/estimated_length=unknown
gap 62037. .62136
/estimated_length=unknown
gap 65660. .65759
/estimated_length=unknown
gap 68285. .68384
/estimated_length=unknown
gap 70706. .70805
/estimated_length=unknown
gap 73032. .73131
/estimated_length=unknown

Query Match 83.2% Score 20.8; DB 14; Length 100161;
Best Local Similarity 91.7%; Pred. No. 7.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 53613 AAAAAAAAACTTCATCATTTAA 53636

RESULT 34
AP007393 101261 bp DNA linear HTG 28-DEC-2004
LOCUS Lot9 corniculatus var. japonicus chromosome 1 clone LJ733H07, ***
DEFINITION SEQUENCING IN PROGRESS ***, 28 unordered pieces.
ACCESSION AP007393
VERSION AP007393.1 GI:56805709
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Lot9 corniculatus var. japonicus (Lotus japonicus)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.

REFERENCE 1 Kaneko, T., Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
AUTHORS Structural Analysis of a Lotus japonicus Genome. XI. Sequence
TITLE Features and Mapping of Nine hundred twenty-one TAC Clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 101261)
AUTHORS Sato, S.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
Department of Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu,
Chiba, 292-0818, Japan (E-mail: ssato@kazusa.or.jp,
URL: http://www.kazusa.or.jp/, Tel: 81-438-52-3935 (ex.2337),
Fax: 81-438-52-3934)

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3172: contig of 3172 bp in length
* 3173 3272: gap of unknown length
* 3273 4648: contig of 1376 bp in length
* 4649 4748: gap of unknown length
* 4749 6568: contig of 1820 bp in length
* 6569 6668: gap of unknown length
* 6669 8116: contig of 1448 bp in length
* 8117 8216: gap of unknown length
* 8217 9822: contig of 1606 bp in length
* 9823 9922: gap of unknown length
* 9923 11529: contig of 1607 bp in length
* 11530 11629: gap of unknown length
* 11630 13431: contig of 1802 bp in length
* 13432 13531: gap of unknown length
* 13532 14732: contig of 1201 bp in length
* 14733 14832: gap of unknown length
* 14833 16318: contig of 1486 bp in length
* 16319 16418: gap of unknown length
* 16419 18296: contig of 1878 bp in length
* 18297 18396: gap of unknown length
* 18397 20467: contig of 2071 bp in length
* 20468 20567: gap of unknown length
* 20568 23188: contig of 2621 bp in length
* 23189 23288: gap of unknown length
* 23289 26553: contig of 3365 bp in length
* 26554 26753: gap of unknown length
* 26754 29267: contig of 2514 bp in length
* 29268 29367: gap of unknown length
* 29368 33864: contig of 4497 bp in length
* 33865 33964: gap of unknown length
* 33965 36461: contig of 2497 bp in length
* 36462 36561: gap of unknown length
* 36562 40463: contig of 3902 bp in length
* 40464 40563: gap of unknown length
* 40564 44838: contig of 4275 bp in length
* 44839 44938: gap of unknown length
* 44939 48249: contig of 3311 bp in length
* 48250 48349: gap of unknown length
* 48350 53768: contig of 5419 bp in length
* 53769 53868: gap of unknown length
* 53869 58661: contig of 4793 bp in length
* 58662 58761: gap of unknown length
* 58762 64480: contig of 5719 bp in length
* 64481 64580: gap of unknown length
* 64581 69403: contig of 4823 bp in length
* 69404 69503: gap of unknown length
* 69504 75189: contig of 5686 bp in length
* 75190 75289: gap of unknown length
* 75290 81168: contig of 5879 bp in length
* 81169 81268: gap of unknown length
* 81269 87661: contig of 6393 bp in length
* 87662 87761: gap of unknown length
* 87762 89459: contig of 11638 bp in length
* 89460 99559: gap of unknown length
* 99560 101261: contig of 1702 bp in length.
Location/Qualifiers
1. .101261
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"
/chromosome="1"
/clone_lib="LJ733H07"
/note="TAC clone:TM0498, synonym:Lotus japonicus"

```

gap      3173..3272
          /estimated_length=unknown
gap      4649..4748
          /estimated_length=unknown
gap      6569..6668
          /estimated_length=unknown
gap      8117..8216
          /estimated_length=unknown
gap      9823..9922
          /estimated_length=unknown
gap      11530..11629
          /estimated_length=unknown
gap      13432..13531
          /estimated_length=unknown
gap      14733..14832
          /estimated_length=unknown
gap      16319..16418
          /estimated_length=unknown
gap      18297..18396
          /estimated_length=unknown
gap      20468..20567
          /estimated_length=unknown
gap      23189..23288
          /estimated_length=unknown
gap      26654..26753
          /estimated_length=unknown
gap      29268..29367
          /estimated_length=unknown
gap      33865..33964
          /estimated_length=unknown
gap      36462..36561
          /estimated_length=unknown
gap      40464..40563
          /estimated_length=unknown
gap      44839..44938
          /estimated_length=unknown
gap      48250..48349
          /estimated_length=unknown
gap      53769..53868
          /estimated_length=unknown
gap      58662..58761
          /estimated_length=unknown
gap      64481..64580
          /estimated_length=unknown
gap      69404..69503
          /estimated_length=unknown
gap      75190..75289
          /estimated_length=unknown
gap      81169..81268
          /estimated_length=unknown
gap      87662..87761
          /estimated_length=unknown
gap      99460..99559
          /estimated_length=unknown
ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 101261;
Best Local Similarity 91.7%; Pred. No. 7.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 AAAAAAAAAACCTTCATCATTTAA 24
DB      54275 AAAAAAAAAACCTTCATTTAA 54298

```

```

RESULT 35
AC109967 0/c
WPCOMBXT
Sequence split into 4 fragments LOCUS AC109967 Accession AC109967
Fragment Name      Begin      End
AC109967_0         1      110000
AC109967_1         100001 210000
AC109967_2        200001 310000

```

```

AC109967.3      300001      378165      DNA      linear      HTG 11-OCT-2002
AC109967      378165 bp
Rattus norvegicus clone CH230-321A19, *** SEQUENCING IN PROGRESS
***, 19 unordered pieces.
AC109967.4      GI:23609307
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 378165)
REFERENCE
AUTHORS
Muzny,D,Marle,M,McIker,M,Lee,A,Abrahamson,S,Adams,C,Alder,J,
Allen,C,Allen,H,Aisbroke,S,Amin,A,Angiano,D,
Ayalebechi,V,Aoyagi,A,Ayodeji,M,Baca,E,Baden,H,
Baldwin,D,Bandaranaike,D,Barber,M,Barnstead,M,Benahmed,F,
Biswal,K,Blair,J,Blankenburg,K,Blych,P,Brown,M,
Bryant,N,Buhay,C,Burch,P,Burrell,K,Calderson,E,
Cardenas,V,Carter,K,Cavazos,I,Ceasar,H,Center,A,
Chacko,J,Chavez,D,Chen,G,Chen,R,Y,Chen,Z,Chu,J,
Cleveland,C,Cockrell,R,Cox,C,Coyle,M,Cree,A,D'Souza,L,
Devita,M,L,Davis,C,Davy-Carroll,L,De Anda,C,Dedrich,D,
Delgado,O,Denson,S,Detamo,C,Ding,Y,Dinh,H,Divya,K,
Draper,H,Dugan-Rocha,S,Dunn,A,Durbin,K,Duval,B,Eaves,K,
Egan,A,Escotto,M,Eugene,C,Evans,C,A,Faller,T,Fan,G,
Fernandez,S,Finley,M,Flagg,N,Forbes,L,Foster,M,Foster,P,
Fraser,C,M,Gabisi,A,Ganta,R,Garcia,A,Garner,T,Garza,M,
Gebregeorgie,E,Geer,K,Gill,R,Grady,M,Guerra,M,Guevara,W,
Gunnaratne,P,Haaland,W,Hamil,C,Hamilton,C,Hamilton,K,
Harvey,Y,Havlak,P,Hawes,A,Henderson,N,Hernandez,J,
Hernandez,R,Hines,S,Hladun,S,L,Hodgson,A,Hogues,M,
Hollins,B,Howell,S,Huylk,S,Hume,J,Idebird,D,Jackson,A,
Jackson,L,Jacob,L,Jiang,H,Johnson,B,Johnson,R,Joliver,A,
Karpach,S,Kelly,S,Kelly,S,Khan,Z,King,L,Kovar,C,
Kowis,C,Kraft,C,L,Ledow,H,Levan,J,Lewis,L,Li,Z,Liu,J,
Liu,Y,Liu,W,Liu,Y,London,P,Longacre,S,Lopez,J,
Lorensuhewa,L,Louisege,H,Lozada,R,J,Lu,X,Me,J,
Maheshwari,M,Mahindartine,M,Mahmoud,M,Mallory,K,Mangum,A,
Mangum,B,Mapua,P,Martin,K,Martin,R,Martinez,E,
Mawhney,S,McLeod,M,P,McNeill,T,Z,Meenen,E,
Mlisaavljevic,A,Miner,G,Mitiga,E,Montemayor,J,Moore,S,
Morgan,M,Morris,K,Morris,S,Munidas,M,Murphy,M,Nair,L,
Nankervis,C,Neal,D,Newton,N,Nguyen,N,Norris,S,
Nwaokemelehu,O,Okwuonu,G,Olanunsgoon,A,Pal,S,Parks,K,
Pasternak,S,Paul,H,Perez,A,Perez,L,Pfankoch,C,
Plopper,F,Polndexter,A,Popovic,D,Primus,E,Pu,L,L,
Piazo,M,Quiroz,V,Rachlin,E,Reeves,K,Regier,M,A,Reigh,R,
Reilly,B,Reilly,M,Ren,Y,Reuter,M,Richards,S,Riggs,F,
Rives,C,Rodkey,T,Rojas,A,Rose,M,Rose,R,Ruiz,S,J,
Sanders,W,Savery,G,Scherer,S,Scott,G,Shatsman,S,Shen,H,
Shetty,J,Shvartbeyn,A,Sison,I,Sitter,C,D,Smaj,S,
Sneed,A,Sodergren,E,Song,X,Z,Socelle,R,Sosa,J,
Steinle,M,Strong,R,Sutton,A,Svatek,A,Tabor,P,Taylor,C,
Taylor,T,Thomas,N,Thomas,S,Tingey,A,Trejos,Z,Usmani,K,
Valas,R,Vera,V,Villasana,D,Waldron,L,Walker,B,Wang,J,
Wang,Q,Wang,S,Warren,J,Warczyk,R,Wooden,H,Worley,K,
Williams,G,Willson,R,Wleczyk,R,Wooden,H,Worley,K,
Wright,D,Wright,R,Wu,J,Yakub,S,Yen,J,Yoon,L,Yoon,V,
Yu,F,Zhang,J,Zhou,X,Zhou,X,Zhou,S,Dunn,D,von
Miederkhausen,A,Weiss,R,Smith,D,R,Holt,R,A,Smith,H,O,
Weinstock,G, and Gibbs,R,A.
Direct Submission
Unpublished
2 (bases 1 to 378165)
REFERENCE
AUTHORS
Worley,K,C.
Direct Submission
Submitted (09-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 378165)
REFERENCE
AUTHORS
Rat Genome Sequencing Consortium.
Direct Submission

```

JOURNAL

COMMENT

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 9, 2002 this sequence version replaced gi:21738214.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GOLT

Center clone name: CH230-321A19

Assembly program: Phrap; version 0.990329

Consensus quality: 278197 bases at least Q40

Consensus quality: 288079 bases at least Q30

Consensus quality: 293331 bases at least Q20

Estimated insert size: 304877; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

NOTE: This sequence may represent more than one clone.

NOTE: This is a 'working draft' sequence. It currently

consists of 19 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 5109: contig of 5109 bp in length

5110 5209: gap of unknown length

5210 170564: contig of 165355 bp in length

170565 170664: gap of unknown length

170665 175608: contig of 4944 bp in length

175609 175708: gap of unknown length

175709 194025: contig of 18317 bp in length

194026 206482: gap of unknown length

206483 206582: contig of 12357 bp in length

206583 217274: gap of unknown length

217275 217374: contig of 10692 bp in length

217375 293004: gap of unknown length

293005 293104: contig of 75630 bp in length

293105 296602: gap of unknown length

296603 296702: contig of 3498 bp in length

296703 303128: gap of unknown length

303129 303228: contig of 6426 bp in length

303229 325004: gap of unknown length

325005 325104: contig of 21776 bp in length

325105 328567: gap of unknown length

328568 328667: contig of 3463 bp in length

328668 333433: gap of unknown length

333434 333533: contig of 4766 bp in length

333534 345832: gap of unknown length

345833 345932: contig of 12299 bp in length

345933 347613: gap of unknown length

347614 347713: contig of 1681 bp in length

347714 348836: gap of unknown length

348837 348936: contig of 1123 bp in length

348937 351223: contig of 2287 bp in length

* 351224 351323: gap of unknown length

* 351324 354288: contig of 2965 bp in length

* 354289 354388: gap of unknown length

* 354389 354566: contig of 10178 bp in length

* 354567 364667: gap of unknown length

* 364667 378165: contig of 13499 bp in length.

Location/Qualifiers

source

1. 378165

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-321A19"

1. 2123

/note="wgs_end_extension"

clone_end:Sp6"

2174. 3258

/note="wgs_end_extension"

clone_end:Sp6"

3409. 5109

/note="wgs_end_extension"

clone_end:Sp6"

5110. 5209

/estimated_length=unknown

9782. 10179

/note="clone_boundary"

clone_end:Sp6"

site:MboI

end sequence:RXANI10TV"

10967. 11776

/note="clone_boundary"

clone_end:T7"

site:MboI

end sequence:RXANI10TV"

150760. 152426

/note="wgs_end_extension"

clone_end:T7"

157563. 159531

/note="wgs_end_extension"

clone_end:T7"

170565. 170664

/estimated_length=unknown

175609. 175708

/estimated_length=unknown

194026. 194125

/estimated_length=unknown

194126. 195876

/note="wgs_end_extension"

misc_feature

Query Match 83.2%; Score 20.8; DB 14; Length 110000;

Best Local Similarity 91.7%; Pred. No. 7.2e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 2 AAAAAAAAACTTCATCTTTTAAAC 25

Db 83351 AAAAAAAAAATTCATCTTTTAAAC 83328

RESULT 36

AL133544/c

LOCUS

DEFINITION

Human DNA sequence from clone R1-71N10 on chromosome 6 contains

the 5' end of a gene for a novel protein, the 5' end of a novel

gene and a Cpg island, complete sequence.

ACCESSION

AL133544.6 GI:7406719

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominiidae; Homo.

REFERENCE

1 (bases 1 to 112085)

AUTHORS
TITLE
JOURNAL

COMMENT

Williams,S.
Direct Submission
Submitted (13-May-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Apr 3, 2000 this sequence version replaced gi:7228194.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest.
RP1-7N10 is from the library RPCL-1 constructed by a YAC.
Pieter de Jong. For further details see
http://www.choxi.org/bacpac/home.htm
VECTOR: pCYPAC2.

FEATURES

source

Location/Qualifiers
1..112085
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RDP:RPCLP704N1071"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP1-7N10"
/clone_lib="RPCL-1"
100
misc_feature
/note="Clone_Right end: RP3-388E23"
join(complement(92627..92653),complement(91022..91146),
complement(67980..68033),complement(66213..66772),
complement(63524..63705),complement(57893..58112),
complement(56133..56325),complement(53740..53835),
complement(48689..48874),complement(47407..47559),
complement(42981..43113),complement(38774..38897),
complement(33426..33655),complement(31607..31713),
complement(30281..30399),complement(29028..29158),
complement(27566..27706),complement(11747..11943),
complement(5350..5376),
complement(AL049552.20:93903..94023),
complement(AL049552.20:57259..57314),
complement(AL049552.20:22289..22451),
complement(AL049552.20:17646..17743),
complement(AL023693.25:187814..187872),
complement(AL023693.25:171300..172961))
/locus_tag="RP1-32B1.2-001"
join(complement(92627..92653),complement(91022..91146),
complement(67980..68033),complement(66213..66772),
complement(63524..63705),complement(57893..58112),
complement(56133..56325),complement(53740..53835),
complement(48689..48874),complement(47407..47559),
complement(42981..43113),complement(38774..38897),
complement(33426..33655),complement(31607..31713),
complement(30281..30399),complement(29028..29158),
complement(27566..27706),complement(11747..11943),
complement(5350..5376),
complement(AL049552.20:93903..94023),
complement(AL049552.20:57259..57314),
complement(AL049552.20:22289..22451),
complement(AL049552.20:17646..17743),
complement(AL023693.25:187814..187872),
complement(AL023693.25:171300..172961))

CDS

complement(AL049552.20:17646..17743),
complement(AL023693.25:187814..187872),
complement(AL023693.25:171300..172961))
/locus_tag="RP1-32B1.2-001"
/note="match: ESTs: A1669190 A1640212 AL048804 A0116951
A0118462 AV726137 AV752371 A0179253 AW236604 A0365646
AW964630 AW967071 B504444 B5671310 B5680572 B5949883
B1653601 BM504518 BM504842 BM704152 BM713796 H29349
match: CDNA: AL136797 BC015019"
join(complement(92627..92653),complement(91022..91146),
complement(67980..68033),complement(66213..66772),
complement(63524..63705),complement(57893..58112),
complement(56133..56325),complement(53740..53835),
complement(48689..48874),complement(47407..47559),
complement(42981..43113),complement(38774..38897),
complement(33426..33655),complement(31607..31713),
complement(30281..30399),complement(29028..29158),
complement(27566..27706),complement(11747..11943),
complement(5350..5376),
complement(AL049552.20:93903..94023),
complement(AL049552.20:57259..57314),
complement(AL049552.20:22289..22451),
complement(AL049552.20:17646..17743),
complement(AL023693.25:187814..187872),
complement(AL023693.25:172940..172961))
/locus_tag="RP1-32B1.2-001"
/standard_name="OTTHUMP0000017263"
/note="match: proteins: Q9H0H2 Q9H803 Q9NXT1"
/codon_start=1
/protein_id="CA12523.1"
/db_xref="GI:56204214"
/db_xref="InterPro:IPR001452"
/db_xref="InterPro:IPR001680"
/db_xref="InterPro:IPR011511"
/db_xref="UniProt/TREMBL:O5TCE9"
/translation="MPFAESBAKKTATVRBELTKTSDLMREKKLKKLVRESENI
SPDITRSNLHYKKESTISDDPTIRSNLPHIKETYSVDSANTNNLKKSTVTKR
NTOLATERNPAGDAVEEDKQKPKKKYIKVPPOLTTDDLKRETEBNKSTHOKTHTK
PQGVQDKSEKANEKREEDTLEBDELMQAYQCHVEEMAKELIKRIKTLKQDLY
FSPDTLPDKLSEKRRKKKEVFEKAEISTLTIGDYEGOKKSSSVSYSDS
HODELTSMSOSTSMODDTPKPKTKKKTKKTKKTVADNHVDVDGHEITRSRPVY
PKCLADDDVLGVYIHTDPLKSPFMSHKVYKHVYDEHNGQYKKKODSRPSSVY
EKENVDYILPTMOPDFOKLSRPLPWEEDYVNEVPYLRSDSPKYLTFEIL
DPLSVDELTKNSSEVQNECGFRKIAMFLKLGANGANSINSLKLOLYPPTPRSP
LSVVEAFEMSKCEPRNHPSTLYTVYVGLVYDICKSYSSMALOEKGPVCRERH
HESSVYTESEKSEKEVIMKRLPGQACIPNGLSLNAGEGCPCLDSHGRIIL
AACASRDGVPILLYEISGRFMREICGHNIITYDLSMSODHYILTSSSGTARIWK
NEINNTTPRVLPKRSVYAKKHPAVRELIVGCVSMIRIKVYKREBDAIIVROP
DVHKSFTNSLCPTDEGHMYSGDCTGVIVVWNTYKINDLESHVHWITNKEIKTER
KGPISLYIEHPNGKRLIHTKOSTLIMDLRIIVAKFVGAAYRKRINSTLPCTG
FLVGSBDGIYVWNPETGQVAMYS8LPKPSJRDISEYHFFEMVAFAGNEPIL
LYYDFVAAOEEAMKPRGQVFPPLPIHOSDMLCPLPHQGSFOIDFVHTES
STWMLVYKRIETVTVIRSCAKVKNVNSFTSPSDELTIKHGDITRPFKQNEH
FGTQGTQITIEKPCNHOVDYATVALYDIYRANRDELTIRHGDITRPFKQNEH
WYSIGIGQGEYFPANHVASETLVQELPPRIKESPLSPBEKTKESPPQKQIN
KJNSQDRIGSESWTKELKKS"
join(complement(56133..56176),complement(53740..53835),
complement(48689..48874),complement(47407..47559),
complement(42981..43113),complement(38774..38897),
complement(33426..33655),complement(31607..31713),
complement(30281..30399),complement(29028..29158),
complement(27566..27706),complement(11747..11943),
complement(5350..5376),
complement(AL049552.20:93903..94023),
complement(AL049552.20:57259..57314),
complement(AL049552.20:22289..22451),
complement(AL049552.20:17646..17743),
complement(AL049552.20:1839..1944),
complement(AL049552.20:535..666),
complement(AL023693.25:187814..187872),
complement(AL023693.25:177737..177839),
complement(AL023693.25:173113..173862),
complement(AL023693.25:172210..172961))

CDs

locus_tag="RP1-32B1.2-002"
 join(complement(56133..56176),complement(53740..53835),
 complement(48689..48874),complement(47407..47559),
 complement(42981..43113),complement(38774..38897),
 complement(33426..33655),complement(31607..31713),
 complement(30281..30399),complement(29028..29158),
 complement(27566..27706),complement(11747..11943),
 complement(5350..5376),
 complement(AL049552.20:39303..94023),
 complement(AL049552.20:57259..57314),
 complement(AL049552.20:22289..22451),
 complement(AL049552.20:17737..17787),
 complement(AL049552.20:1839..1944),
 complement(AL023693.25:187814..187872),
 complement(AL023693.25:17737..17787),
 complement(AL023693.25:173713..173862),
 complement(AL023693.25:172210..172961),
 locus_tag="RP1-32B1.2-002"
 /note="match: ESTs: A1653410 A1808115 A1954664 AV726137
 AW103846 BG181752 BG214521
 match: CDNAS: AK000076 AK024085"
 join(complement(47407..47538),complement(42981..43113),
 complement(38774..38897),complement(33426..33655),
 complement(31607..31713),complement(30281..30399),
 complement(29028..29158),complement(27566..27706),
 complement(11747..11943),complement(5350..5376),
 complement(AL049552.20:39303..94023),
 complement(AL049552.20:57259..57314),
 complement(AL049552.20:22289..22451),
 complement(AL049552.20:17737..17787),
 complement(AL049552.20:17646..17743),
 complement(AL049552.20:1936..1944),
 locus_tag="RP1-32B1.2-002"
 /standard_name="OTTHUMP0000017264"
 /codon_start=1
 /protein_id="CA12522.1"
 /db_xref="GI:56204213"
 /db_xref="InterPro:IPR001452"
 /db_xref="InterPro:IPR001680"
 /db_xref="InterPro:IPR011511"
 /db_xref="UniProt/TrEMBL:Q5TF10"

Query Match 83.2%; Score 20.8; DB 8; Length 112085;
 Best Local Similarity 91.7%; Pred. No. 7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTTCATCATTTAA 24
 |||||
 Db 72076 AAAAAAAAACTTCATTTAA 72053

RESULT 37
 AC020202 118591 bp DNA linear HTG 03-JAN-2000
 LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 ACCESSION AC020202.1 GI:6664695
 VERSION HTG, HTGS PHASE2.
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydrioidae; Drosophilidae; Drosophila.
 1 (bases 1 to 118591)
 Adams, M. and Venter, J.C.
 REFERENCE Direct Submission
 TITLE Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
 JOURNAL Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10212386 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES
 Location/Qualifiers
 1..118591
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 118591;
 Best Local Similarity 91.7%; Pred. No. 7e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTTCATCATTTAA 24
 |||||
 Db 92616 AAAAAAAAACTTCATTTAA 92639

RESULT 38
 HS1158B12 122884 bp DNA linear PRI 18-MAY-2005
 LOCUS Human DNA sequence from clone RP5-1158B12 on chromosome
 DEFINITION Xp11.21-11.4 Contains the ZXDA gene for X-linked duplicated zinc
 finger A, a v-myc myelocytomatosis viral oncogene homolog 1 lung
 carcinoma derived (avian) (MYCL1) pseudogene, a keratin 8 (KR18)
 pseudogene and a CpG island, complete sequence.
 AL034396
 AL034396.6 GI:4902597
 HTG: keratin; KR18; MYCL1; ZXDA.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euteheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 1 (bases 1 to 122884)
 Bird, C.
 REFERENCE Direct Submission
 TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
 JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
 COMMENT Clone requests: clonerequest@sanger.ac.uk
 On May 27, 1999 this sequence version replaced gi:4678710.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/ChrX
 RP5-1158B12 is from the library RP1-5 constructed by the group of
 Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCYPAC2
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: vegas@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one subclone; and the assembly was confirmed by restriction digest,
 except on the rare occasion of the clone being a YAC.
 Location/Qualifiers
 1..122884
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="RZPD:RP1704B121158"
 /db_xref="taxon:9606"
 /map="p11.21-11.4"

	gene	/clone="RP5-1158B12" /clone_1ib="RBCI-5" complement(19812..23924) /gene="ZXDA" /locus_tag="RP5-1158B12.3-001" /locus_tag="RP5-1158B12.3-001" complement(19812..23924) /gene="ZXDA" /locus_tag="RP5-1158B12.3-001" /product="zinc finger, X-linked /note="match: ESTs: A1766017.1 BM679833.1 BM829497.1 CB988499.1 match: CDNAS: L14787.1 L14788.1" complement(19812)
	polyA_site	/gene="ZXDA" /locus_tag="RP5-1158B12.3-001" complement(19835) /gene="ZXDA" /locus_tag="RP5-1158B12.3-001" complement(21125..23524) /gene="ZXDA"
	CDS	/locus_tag="RP5-1158B12.3-001" /standard_name="OTTHUMP00000023422" /note="match: proteins: P98168 P98169" /codon_start=1 /product="zinc finger, X-linked, duplicated A" /protein_id="CAB46717.1" /db_xref="gi:544319" /db_xref="Genew:13198" /db_xref="GOA:P98168" /db_xref="HSP:1SP2" /db_xref="InterPro:IPR007087" /db_xref="UniProt/Swiss-Pro: P98168" translation="MEIKPLPARGTLQGGGAGGI PACGGRHKGPDPACGVPTRRIL LPFGPDGCGRRREASTASRGSGELFAPRHPSGCGDDPFVLVLDVGDEVERE AGSOAAAPVLRBEAKAGPGLOGBESGNPNPGSCAQPHCLSAVTPRAISAPAHAMA FAGVTIHNODLLRFENGVTLTATPRHPAPRAAQAQRCGLIARQGRPAQANNA DCPELRSDILAEPARPAPARAPOBEAEGLAALAGPROLDSGEVAVITCEPALCGE TEFAKHDLQKMLLTHSSSQGQRPCPKIGGCGWFTTSYKLRHLQSHDKLRPFCGCP EGCKSTFTYNLKAHKMGHEONSFKCEVEBSFPDARKGAHQRSHEPERYOCASF FSGCKGTFTYSALFESHRAHFREQLEFSCSFPGCSKOYDACRLKIHRHTGERP LCDDGCGMTFTMSKLIRHKRKDDDRFPCPVGCGSKPTRABHLKGHSITHLGTRP PFCVPVAGCCAFARSLSLYTHSKKHODVDTMRCPISSCNLFYSKSMKTHMYK RAHYGQDILADLEANSLTPSESELTSORONDLSAEIVLSFDVDPDSTALUDTAL NSGILLTIDVASVSTLAGHLPPANNNSVGAOVDPDSLWATSDPQSIDTSLFETAAV GFQOSSLTMDEVSSVGPLGSLDSLAMKNSSPPOLATPSSKLTVDITLTPESTYL ENSVELTTPAKAEWSVHPNSDFPGQEGETQFGPMAAGNHGSQKERNLIVTSSSFL V"
	gene	complement(46941..47907) /locus_tag="RP5-1158B12.2-001" /pseudo complement(46941..47907) /locus_tag="RP5-1158B12.2-001" /note="match: proteins: O43372 P10166 P12524 P12525 Q14697" /pseudo /codon_start=1 /product="v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian) (MYC1L) pseudogene" 97790..99237 /locus_tag="RP5-1158B12.1-001" /pseudo 97790..99237 /locus_tag="RP5-1158B12.1-001" /locus_tag="RP5-1158B12.1-001" /locus_tag="RP5-1158B12.1-001" /note="match: proteins: O42434 O93532 P02538 P05786 P05787 P08776 P16579 P16878 P18520 P35908 Q01546 Q10758 Q61518 Q91219" /pseudo /codon_start=1 /product="keratin 8 (KRt8) pseudogene"
ORIGIN		
Query Match	83.2%	Score 20.8; DB 8; Length 122884;
Best Local Similarity	91.7%;	Pred. No. 6.9e+02;

Matches	22: Conservative	0	Mismatches	2	Indels	0	Gaps	0
Db	20569	AAAAAAAAACCTTCATCATTTAA	24					
RESULT 39								
LOCUS	AC007837							
DEFINITION	Drosophila melanogaster, chromosome 2R, region 57B2-B3, BAC clone BACR04107, complete sequence.							
ACCESSION	AC007837	123647 bp	DNA	linear	INV 21-FEB-2001			
VERSION	AC007837							
KEYWORDS	AC007837.4	GI:13027506						
SOURCE	HTG.							
ORGANISM	Drosophila melanogaster (fruit fly)							
REFERENCE	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidae; Drosophilidae; Drosophila. 1 (bases 1 to 123647) Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gockayne,J.D., Amanosides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busan,D.A., Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M., Dodson,K., Doreste,V., Doup,L.E., Doyle,C., Dreger,K.D., Farfan,D., Ferreira,S., Frisze,E., Galle,R.F., Gary,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Idegami,C., Jalali,M., Kruse,D., Li,P., Mettel,B., Moshrefi,A., McInerosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacled,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svrtskas,R., Tector,C., Williams,S.M., Zaveri,U.S., Smith,H.O., Rubin,G.M. and Venter,J.C. Sequencing of Drosophila chromosome 2R, region 57B2-B3 Unpublished 2 (bases 1 to 123647) Celniker,S.E., Agbayan,I., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champagne,M., Chavez,C., Chew,M., Cieiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,B., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacled,B., Park,S., Pfeiffer,B., Poon,L., Segueira,A., Sethi,H., Snir,E., Svrtskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M. Direct Submission Submitted (16-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA On Feb 21, 2001 this sequence version replaced gi:5670578. Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdg@fruitfly.berkeley.edu . Location/Qualifiers 1. 123647 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /strain="Y; cn bw sp" /db_xref="taxon:7227" /chromosome="2R" /map="57B2-B3" /clone="BACR04107 (D644)" /clone_1ib="RPCT-98 (Roswell) Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6")							

Query Match 83.2%; Score 20.8; DB 2; Length 123647;
Best Local Similarity 91.7%; Pred. No. 6.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 123554 AAAAAAAAAACTTAATCATTTAA 123577

RESULT 40
BX942828 125793 bp DNA linear VRT 07-JUN-2005
DEFINITION Zebrafish DNA sequence from clone CH211-237F16 in linkage group 17,
complete sequence.
ACCESSION BX942828 GI:67008330
VERSION
KEYWORDS
SOURCE HTG.
ORGANISM Danio rerio (zebrafish)

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (07-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests:
http://www.sanger.ac.uk/Projects/D_rerio/fqgs.shtml#atweight
On Jun 7, 2005 this sequence version replaced gi:66393041.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: zfsh-help@sanger.ac.uk

COMMENT

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Any regions longer than 1kb tagged as misc-feature 'unsure' are part of a tandem repeat of more than 10kb in length where it has not been possible to anchor the base differences between repeat copies. The region has been built up based on the repeat element to match the total size of repeat indicated by restriction digest, but repeat copies may not be in the correct order and the usual finishing criteria may not apply. CH211-237F16 is from a CHORI-211 BAC library
VECTOR: pTARBAC2.1.
Location/Qualifiers
1. 125793

FEATURES
Source
Location/Qualifiers
1. 125793

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/chromosome="17"
/clone="CH211-237F16"
/clone_id="CHORI-211"

ORIGIN

Query Match 83.2%; Score 20.8; DB 5; Length 125793;

Best Local Similarity 91.7%; Pred. No. 6.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 11239 AAAAAAAAAACTTAATCATTTAA 11262

RESULT 41
AC108110 136730 bp DNA linear PRI 23-APR-2002
LOCUS
DEFINITION Homo sapiens chromosome 5 clone RP11-26506, complete sequence.
ACCESSION AC108110 GI:20270109
VERSION
KEYWORDS
SOURCE HTG.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 136730)
DOE Joint Genome Institute.
DIRECT Submission
Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 136730)
DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

REFERENCE
AUTHORS
TITLE
JOURNAL
DIRECT Submission
Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 23, 2002 this sequence version replaced gi:19224882.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 168.7kb). It is clipped at the overlap with AC026444.
The number of bases overlapped is 10049.

VERSION AC007915.3 GI:7243815
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (chale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
AUTHORS
1 (bases 1 to 137336)
Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C.,
Shim, P., Altafi, H., Bei, Q., Chin, C., Chlou, J., Choi, E., Conn, L.,
Conway, A., Gonzalez, A., Hansen, N., Howng, B., Koo, T., Lam, B.,
Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharbek, N.,
Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A.,
Thavert, A., Toriumi, M., Vaysberg, M., Yu, G., Federpiel, N. A.,
Theologis, A. and Ecker, J. R.
Genomic sequence for Arabidopsis thaliana BAC F27F5 from chromosome
I

TITLE
Unpublished
2 (bases 1 to 137336)
Ecker, J. R.
Direct Submission
Submitted (25-JUN-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 137336)
Ecker, J. R.
Direct Submission
Submitted (15-MAR-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 137336)
Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C.,
Shim, P., Altafi, H., Bei, Q., Chin, C., Chlou, J., Choi, E., Conn, L.,
Conway, A., Gonzalez, A., Hansen, N., Howng, B., Koo, T., Lam, B.,
Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharbek, N.,
Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A.,
Thavert, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R.,
Federpiel, N., Theologis, A. and Ecker, J.
Direct Submission
Submitted (10-MAY-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
5 (bases 1 to 137336)
Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C.,
Shim, P., Altafi, H., Bei, Q., Chin, C., Chlou, J., Choi, E., Conn, L.,
Conway, A., Gonzalez, A., Hansen, N., Howng, B., Koo, T., Lam, B.,
Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharbek, N.,
Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A.,
Thavert, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R.,
Federpiel, N., Theologis, A. and Ecker, J.
Direct Submission
Submitted (11-OCT-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Mar 15, 2000 this sequence version replaced gi:5913366.
Location/Qualifiers
1. 137336
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F27F5"
join(<47. .208,291. .416)
/note="similar to MCM2-related protein emb|CAA69609.1"
/codon_start=1
/evidence=not experimental
/product="F27F5.1"
/protein_id="AA069149.1"
/db_xref="GI:7767652"
/translation="SPKRYITFKDPSNLLVLKELVKNALKPEEIIISGNSGLPTI
EKKIEELATKAEYDIADLRPFSTSPSKAHFELDGRGMIKCPKRLITM"
join(1889. .1924,2230. .2547,2637. .2822,3343. .3404,
3604. .3736,4180. .4232,4317. .4374,4453. .4743,5113. .5211,

CDS
5312. .5380,5488. .5565,5739. .5825,5912. .5946,6093. .6141,
6512. .6592,6680. .6772,6839. .6892,6977. .7093,7172. .7270,
7380. .7431,7532. .7593,7736. .7843,7925. .7989,8091. .8193,
8278. .8592,8668. .8778)
/note="unknown protein; similar to EST gb|A197587.1"
/codon_start=1
/evidence=not experimental
/product="F27F5.2"
/protein_id="AA069150.1"
/db_xref="GI:7767653"
/translation="MANPPOSSGCTQGFPMVIRGQOQHGVPAASQGFPHGVHPPV
QSQPPXSQPTQQOQLFPPVRGDPVHITSSQAVSVYITQNKLTISGSTQPNAP
MTGSPQSPSSPYTVTPSSYPOQQTSLVQPSQNVHGVAPPAATWPPVQVST
LSPVQQTGGQTPAVASTDGNLTPOQASDMOEBTSDGKCTLPHGSMNLSYLTIT
YLSRYNNKRTKQSMWEKPLEMTPLRPAASVYWKFTPEGKYYNKKYTESKMT
IPEDKYKARQOALASEKTSLEAGSTPLRSHAASSDLAVSYTVSVSPSSALKTG
SSSPDAGLAVPVRPSPVAPVPTPQSAIDTETATMYPSISGPAKMKSNVCKAN
LSPAGDANVBPVYATKQDAPAKAFKSLBSNVVNSDMTWEOITKXIVHDKYGLAR
TLGRKQAFNEVILGQRKVAEERRRQKAREEVKMLECEBLSLAKMRDEDLF
DNYIVELERKEREKAEERHQAQYADVYKPLETCDYIKAGTQWRKIQRLBEDDCSL
EKIDRLIGPEEYILDLKEEBELKRVKERRARERKRDARFTLLBEHVAAGILYAK
TYMDYCIELKQDLPQYAVASNTSGTPKDLFEDVTELEKQYHEDKSYYVDAKMSK
ANKSAISELSQTSIDINYLKIYDQVYKEREKARKQRLAEPNLLHTPK
ETVYASWESKQLVESQYRSIGDSVQGLPEEYITSLQEPKAKERKQDEKRA
KERERDEKERKQDEKEREKEREKEREKEREKEREKEREKEREKEREKEREK
KDRDRKRRRHNNNSDEVDSDRDRDBESKSSRKQNDKRSKHANSPESEENRH
KQKESRSRSGMDELDEGEV"
complement(join(9246. .9284,9349. .9534,9627. .9739,
9825. .9952,10049. .10368))
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="F27F5.3"
/protein_id="AA069175.1"
/db_xref="GI:7767678"
/translation="MRMLVLRTPPGVSAVRPITTKKSPSSPSSDLSLLPAPNTHRSR
PRVVALVAGDSVNSKPSASQIPSSGDETVFGQDNVPLEVIOFDKSSYSSSN
ITKWRVALLAGDGVALLIFSAIGRSPFVPSLDTLHTADPFIPLGSAVFLG
YAEGRMGKSKRAVNAAKSWITVGLGIIISVASSGHIPIYFVFLTMGSTAVLLI
GMALLPSVLPSTSKKDDPYRKSGAELEPEYIPLCKGPAE"
join(11070. .11093,11262. .11372,11448. .11477,11558. .11629)
/note="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="F27F5.4"
/protein_id="AA069151.1"
/db_xref="GI:7767654"
/translation="WVGQIVISGCCMRKRSKGTGMSGLAVETISARFVGSVARCAE
ARKLVLSAMEVSGSDVGFGLVARIIDNGHVARV"
join(25536. .25840,26927. .27403,27965. .28045)
/note="unknown protein; similar to EST gb|T44906.1"
/codon_start=1
/evidence=not experimental
/product="F27F5.5"
/protein_id="AA069152.1"
/db_xref="GI:7767655"
/translation="MKKMLIAATVAITAIIVRQGTQYQGNKBALEKLEWSDRLG
IMAIPTVAVHTITLALCPHAVPEPAGSLMGLPALCVPSAKYLAASFWIR
FVFKSSRAAGAHSNKYPNILRQVGRDQKVVILARPEISYVINYVLAAREVAF
VADPLPVTIVIGLPMIIQNASVGSILAMAAVASAGKQSGVWGVPFVLGLISVLLS
LRIKKYSAGITEASDTSANSSIASSETNPAYGVGDSKKSEHMSFLBLVOGLLD
NNEV"
join(28197. .28466,28552. .28743,28830. .28995,29111. .29523)
/note="peroxidase-like protein emb|CAB6621.1; similar to
ESTs gb|T76544.1, gb|AA394647.1"
/codon_start=1
/evidence=not experimental
/product="F27F5.6"
/protein_id="AA069153.1"
/db_xref="GI:7767656"
/translation="MAISKILPVLVLPVLPSPVSVVAPHLGFGWGSNPSPIGFSYFN
LYPQFPQSGPQADELYMTVTELEAIKEPRMAASLRHLHHDPCFVQSCDASITLDDSA
TIRSEKNAGFNKNSVRGPOVIDEIKAKLEQACQVQSCADITLALARGSTILSGPSW

[illegible]

	AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 141674)	
AUTHORS	DOE Joint Genome Institute.	
TITLE	Direct Submission	
JOURNAL	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
REFERENCE	3 (bases 1 to 141674)	
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.	
TITLE	Direct Submission	
JOURNAL	Submitted (22-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
REFERENCE	4 (bases 1 to 141674)	
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.	
TITLE	Direct Submission	
JOURNAL	Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
COMMENT	On Feb 22, 2000 this sequence version replaced gi:6165068. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing completed at Stanford Human Genome Center www.hgsc.stanford.edu Quality: phrap Quality >=40 99.8% of Sequence; Estimated Total Number of Errors is 0.4.	
FEATURES	Location/Qualifiers	
source	1..141674 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="5" /clone="CTD-2035K4"	
ORIGIN		
Query Match	83.2%; Score 20.8; DB 8; Length 141674;	
Best Local Similarity	91.7%; Prod. No. 6.6e+02;	
Matches	22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	1 AAAAAAAAAACCTTCATCATTTTAA 24	
Ddb	8262 ATAAAAAACCCCTTCATCATTTTAA 8285	
RESULT 44		
AC162580/c		
LOCUS	AC162580	
DEFINITION	Bos taurus clone CH240-111ID18, *** SEQUENCING IN PROGRESS ***, 14 unordered pieces.	
ACCESSION	AC162580	
VERSION	AC162580.3 GI:68101957	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.	
SOURCE	Bos taurus (cow)	
ORGANISM	Bos taurus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.	
AUTHORS	1 (bases 1 to 144964) Muzny,D.,Marie., Metzker,M.,Lee, Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooke,S., Amin,A., Angiano,D., Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaruaike,D., Barber,M., Barnstead,M., Benahmed,F., Bielawski,K., Blair,J., Blankenburg,K., Blythe,P., Brown,K., Bryant,N., Bukey,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A., Chacko,U., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Derich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Gan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabriel,A., Gante,R., Garcia,A., Garner,T., Garza,M., Gebreyes,E., Geier,K., Gill,R., Grady,M., Guerra,W., Guenard,W.,	

Gunaratne, P., Healand, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollander, B., Howells, S., Huljk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowalski, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangun, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nair, T., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokilemeh, O., Okwou, G., Olarnpungoon, A., Pal, S., Parke, K., Paternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L. L., Puzo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wiczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Unpublished
2 (bases 1 to 14964)
Worley, K. C.
Direct Submission
Submitted (30-MAY-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 14964)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2005 this sequence version replaced gi:66796318.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: FRFB
Center clone name: CH240-111018
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 138998 bases at least Q40
Consensus quality: 140732 bases at least Q30
Consensus quality: 142106 bases at least Q20

Estimated insert size: 144890; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
1

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
12182: contig of 12182 bp in length
12183
12438: gap of 255 bp
12438
14131: contig of 1694 bp in length
14132
14181: gap of 50 bp in length
14182
20628: contig of 6447 bp in length
20629
20767: gap of 139 bp
20768
24248: contig of 3481 bp in length
24249
24298: gap of 50 bp
31465
31466: contig of 7168 bp in length
31467
31738: gap of 272 bp in length
31739
31759: contig of 42964 bp in length
31759
74702: gap of 50 bp
74703
74752: gap of 50 bp
74753
81414: contig of 6662 bp in length
81415
81465: gap of 50 bp
81465
81465: contig of 1701 bp in length
81465
83233: gap of 68 bp
83234
83234
85820: contig of 2587 bp in length
85821
85821: gap of 50 bp
85821
93719: contig of 7849 bp in length
93720
93720: gap of 50 bp
93720
97106: contig of 3337 bp in length
97107
97628: gap of 522 bp in length
97629
103748: contig of 6121 bp in length
103750
103759: gap of 50 bp
103800
142777: contig of 38978 bp in length
142778
142964: gap of 187 bp
142965
144964: contig of 2000 bp in length.
* * *
Location/Qualifiers
1. 14964
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-111D18"
12183. 12437
/estimated_length=255
14132. 14181
/estimated_length=50
20629. 20767
/estimated_length=139
24249. 24298
/estimated_length=50
31467. 31738
/estimated_length=272
74703. 74752
/estimated_length=50
81415. 81464
/estimated_length=50
83166. 83233
/estimated_length=68
85821. 85870
/estimated_length=50
93720. 93769
/estimated_length=50
97107. 97628
/estimated_length=522
103750. 103799
/estimated_length=50
142778. 142964
/estimated_length=187

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 144964;
 Best Local Similarity 91.7%; Pred. No. 6.6e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTTAA 24
 Db 82745 AAAAAAAAACTTCATCATTTTAA 82722

RESULT 45
 CR548641 146043 bp DNA linear HTG 08-MAY-2005
 LOCUS Danio rerio clone CH211-255N14, WORKING DRAFT SEQUENCE, 6 unordered
 DEFINITION pieces.
 ACCESSION CR548641.2 GI:63094482
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Danio rerio (zebrafish)
 SOURCE Danio rerio
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 146043)
 McLaren, S.
 Direct Submission
 Submitted (07-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On May 8, 2005 this sequence version replaced gi:50057677.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: zc255N14
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Consensus quality: 144306 bases at least Q40
 Consensus quality: 144611 bases at least Q30
 Consensus quality: 144960 bases at least Q20
 Insert size: 145543; sum-of-coverage
 Insert size: 145810; 3.6% error; agarose-gel
 Quality coverage: 9.92x in Q20 bases; sum-of-coverage
 Quality coverage: 9.91x in Q20 bases; agarose-gel

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 9146: contig of 9146 bp in length
 * 9147 9246: gap of 100 bp
 * 9247 60405: contig of 51159 bp in length
 * 60406 60505: gap of 100 bp
 * 60506 92241: contig of 31736 bp in length
 * 92242 92341: gap of 100 bp
 * 92342 98312: contig of 5971 bp in length
 * 98313 98412: gap of 100 bp
 * 98413 102852: contig of 4440 bp in length
 * 102853 102952: gap of 100 bp
 * 102953 146043: contig of 43091 bp in length.

FEATURES
 source 1..146043
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-255N14"

misc_feature
 1..9146
 /note="assembly_fragment:00133
 clone_end:T7
 vector_side:left"

misc_feature
 9247..60405
 /note="assembly_fragment:01421"
 60506..92241
 /note="assembly_fragment:00230
 fragment_chain:1"

misc_feature
 92342..98312
 /note="assembly_fragment:00005
 fragment_chain:1"

misc_feature
 98413..102852
 /note="assembly_fragment:00068
 fragment_chain:1"

misc_feature
 102953..146043
 /note="assembly_fragment:00776
 fragment_chain:1
 clone_end:SP6
 vector_side:right"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 146043;
 Best Local Similarity 91.7%; Pred. No. 6.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTTAA 24
 Db 7406 AAAAAAAAACTTCATCATTTTAA 7429

RESULT 46
 BS000106 146196 bp DNA linear PRI 12-JUN-2004
 LOCUS Pan troglodytes chromosome 22 clone:PTB-082114, map 22, complete
 DEFINITION sequences.
 ACCESSION BS000106 BA0000046
 VERSION BS000106.1 GI:37537373
 KEYWORDS HTG.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Pan.

REFERENCE 1
 AUTHORS The International Chimpanzee Chromosome 22 Consortium.
 TITLE DNA sequence and comparative analysis of chimpanzee chromosome 22
 JOURNAL Nature 429, 382-388 (2004)
 AUTHORS Hattori, M., Toyoda, A., Watanabe, H., Taylor, T.D., Kuroki, Y.,
 Fujiyama, A. and Sakaki, Y.
 DIRECT SUBMISSION
 Submitted (12-MAY-2003) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT 2 (bases 1 to 146196)
 *The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
 *Chinese National Human Genome Center at Shanghai, Shanghai, China;
 *GDF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
 of Molecular Biotechnology, Jena, Germany; *KIRB Genome Research
 Center, Daejeon, Korea;
 *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
 *National Institute of Genetics, Mishima, Japan;
 *National Yang Ming University Genome Research Center, Taipei,
 Taiwan;
 *RIKEN Genomic Sciences Center, Yokohama, Japan.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp

----- Project Information -----

```

Center project name: The Chimpanzee Chromosome 22 Sequencing Project
Center Clone name: PRS-082114
----- Summary Statistics -----
Sequencing vector: pUC18, pUC13, pTZ19R; 100% of reads Chemistry:
Dye-terminator Big Dye and ET; 100% of reads Assembly program:
Phrap, version 0.990329
Consensus quality: 147,012 bases at least Q40
Consensus quality: 1,088 bases at least Q30
Consensus quality: 91 bases at least Q20
Consensus quality:

```

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30). An attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Source information: The RPCT-43 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at <http://www.chori.org/bacpac/mchimp3.htm>. The clone may be obtained from Pieter J. de Jong and coworkers (<http://www.chori.org/bacpac>).

The CHORI-251 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa, K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Bacci Zink, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at <http://www.chori.org/bacpac/chimpanzee251.htm>. The clone may be obtained from Pieter J. de Jong and coworkers (<http://www.chori.org/bacpac>).

The PTB1 chimpanzee BAC library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.

Clones may be obtained from Asao Fujiyama and co-workers (<http://www.gsc.riken.go.jp>).

The PTF22 chimpanzee Fosmid library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.

Clones may be obtained from Asao Fujiyama and co-workers
(<http://www.gsc.riken.go.jp>).
VECTOR: pKS143

VECTOR: pKS143

Sequence Quality Assessment:

This entry has been annotated with sequence estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Neighboring clones: RP43-117B17 (left) and RP43-093B21 (right) .

FEATURES

```
1. 148196
   /organism="Pan troglodytes"
   /mol_type="genomic DNA"
   /db_xref="taxon:9598"
   /chromosome="22"
   /clone="PTB-082114"
   /clone_1fb="PTB1 chimpanzee BAC"
```

ORIGIN

	Query Match	83.2%	Score 20.8;	DB 8;	Length 148196;
	Best Local Similarity	91.7%;	Pred. No. 6.5e+02;		
	Matches 22; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1 AAAAAAAAAACCTCATCTTTAA	24			
Db	24939 AAAAAAAAAAGTTCATCTTTAA	24962			

RESULT 47

LOCUS	AC007422	156485 bp	DNA	linear	PRI 09-JUN-2000
DEFINITION	Homo sapiens chromosome , clone RPL-68A1, complete sequence.				
ACCESSION	AC007422				
VERSION	AC007422.7	GI:7801434			
KEYWORDS	HTS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				

REFERENCE

AUTHORS Birten, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens chromosome, clone RP11-68A1
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 156485)
 AUTHORS Birten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Anderson, M.,

TITLE

JOURNAL Submitted (28-APR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 156485)

AUTHORS

Anderson, S., Baldwin, J., Barua, N., Baetien, V., Beda, F., Boguslavsky, L., Bouhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArlelano, K., Dewar, K., Diaz, J., S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardina, S., Ginde, S., Goylete, M., Graham, L., Grand, P., Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karttas, A., Klein, J., LaRocque, K., Lamazeres, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrum, J., Menaus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C., H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tsafaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Vziel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
Direct Submissions
JOURNAL
Submitted (30-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 130 Charles Street, Cambridge, MA 02141, USA
On May 14, 2000 this sequence version replaced gi:7139790.
COMMENT
All repeats were identified using RepeatMasker:

```
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L638
Center clone name: 68_A_1

FEATURES
Source
----- Location/Qualifiers
1..156485
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-68A1"
/clone_lib="RPC1-11 Human Male BAC"

repeat_region
/rpt_family="THE1C"
1..229

repeat_region
/rpt_family="THE1C"
2121..2423

repeat_region
/rpt_family="MER2"
2459..2812

repeat_region
/rpt_family="THE1A"
3298..3532

repeat_region
/rpt_family="THE1C"
complement(3547..5461)

repeat_region
/rpt_family="L1PA2"
5456..6796

repeat_region
/rpt_family="L1PA2"
6798..6943

repeat_region
/rpt_family="THE1C"
7525..7715

repeat_region
/rpt_family="L2"
7936..8042

repeat_region
/rpt_family="Tigger3 (Golem)"
8048..8779

repeat_region
/rpt_family="Tigger3 (Golem)"
9581..9623

repeat_region
/rpt_family="(TAA)n"
11174..11470

repeat_region
/rpt_family="AluDb"
complement(11832..12099)

repeat_region
/rpt_family="MER2"
12602..12902

repeat_region
/rpt_family="AluSp"
16079..16178

repeat_region
/rpt_family="ATRS8"
complement(16488..16916)

repeat_region
/rpt_family="LOR1b"
complement(16963..17249)

repeat_region
/rpt_family="AluOo"
complement(17367..17599)

repeat_region
/rpt_family="MIR"
17997..18019

repeat_region
/rpt_family="(CA)n"
complement(18373..18540)

repeat_region
/rpt_family="LMD3"
20552..20579

repeat_region
/rpt_family="(TTTA)n"
complement(21039..21123)

repeat_region
/rpt_family="L2"
complement(23416..23526)

repeat_region
/rpt_family="MIR"
23628..23660

repeat_region
/rpt_family="AT_rich"
23903..23947

repeat_region
/rpt_family="T_rich"
24302..24342

repeat_region
/rpt_family="AT_rich"
25199..26765

repeat_region
/rpt_family="L1MC1"
26924..27073

repeat_region
/rpt_family="MER6A"
27104..27399

repeat_region
/rpt_family="AluY"
27404..27614

repeat_region
/rpt_family="L1MC1"
complement(27615..27654)

repeat_region
/rpt_family="THE1B"
27655..27880

repeat_region
/rpt_family="AluSg/x"
complement(27881..28174)

repeat_region
/rpt_family="THE1B"
28179..28252

repeat_region
/rpt_family="L1MC1"
28479..29862

repeat_region
/rpt_family="Char11e1a"
complement(30122..30493)

repeat_region
/rpt_family="LTR16C"
complement(30608..30791)

repeat_region
/rpt_family="L1ME3"
31060..31428

repeat_region
/rpt_family="L1ME3A"
31430..31532

repeat_region
/rpt_family="(TA)n"
31648..31804

repeat_region
/rpt_family="(TAGA)n"
32092..32125

repeat_region
/rpt_family="AT_rich"
32157..33391

repeat_region
/rpt_family="L2"
33498..33553

repeat_region
/rpt_family="L2"
33957..34122

repeat_region
/rpt_family="MIR"
complement(34518..34977)

repeat_region
/rpt_family="L1PB1"
34978..35417

repeat_region
/rpt_family="MSTRA"
complement(35418..38924)

repeat_region
/rpt_family="L1PB1"
complement(38925..39224)

repeat_region
/rpt_family="AluOo"
complement(39225..41580)

repeat_region
/rpt_family="L1PB1"
complement(41912..42065)

repeat_region
/rpt_family="L1M2"
complement(42054..42218)

repeat_region
/rpt_family="L1PA15-16"
complement(42164..42839)

repeat_region
/rpt_family="L1PB1"
complement(43252..43648)

repeat_region
/rpt_family="L1PA15-16"
complement(43696..44059)

repeat_region
/rpt_family="THE1B"
complement(44060..45595)

repeat_region
/rpt_family="THE1B-internal"
complement(45596..45959)

repeat_region
/rpt_family="THE1B"
complement(47790..47842)

repeat_region
/rpt_family="L2"
47908..47947

repeat_region
/rpt_family="(TTTTA)n"
complement(47948..48232)

repeat_region
/rpt_family="AluSx"
complement(48475..48676)

repeat_region
/rpt_family="MIR"
complement(49281..49384)

repeat_region
/rpt_family="L2"

Query Match 83.2%; Score 20.8; DB 8; Length 156485;
Best Local Similarity 91.7%; Pred. No. 6.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAACCTTCATCTTAA 24
```


Db 24285 AAAAAAAAACTTATCATTTTAA 24308

|||||

RESULT 48
AC009218 156753 bp DNA linear INV 10-MAR-2001
LOCUS Drosophila melanogaster, chromosome 2R, region 57B-57B, BAC clone
DEFINITION BACR33D17, complete sequence.
ACCESSION AC009218
VERSION AC009218.7 GI:13270559
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Ceiliker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amentides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferrera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalili,M., Kruse,D., Li,P., Matzel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nimmo,J.,
Pacile,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Shuanaenavong,S., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
TITLE Sequencing of Drosophila chromosome 2R, region 57B-57B
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156753)
AUTHORS Ceiliker,S.E., Agbayan,A., Arcaina,T.T., Baxter,B., Blazek,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katta,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Mar 10, 2001 this sequence version replaced gi:6978375.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.
FEATURES
source
1..156753
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="57B-57B"
/clone="BACR33D17 (D945)"
/clone_lib="RPCT-98 (Rosewell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACe3.6)"
ORIGIN
Query Match 83.2%; Score 20.8; DB 2; Length 156753;
Best Local Similarity 91.7%; Pred. No. 6.4e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTATCATTTTAA 24
Db 156660 AAAAAAAAACTTATCATTTTAA 156683

|||||

RESULT 49
AL929072/c 160432 bp DNA linear VRT 30-JAN-2003
LOCUS zebrafish DNA sequence from clone CH211-140B8, complete sequence.
DEFINITION AL929072
ACCESSION AL929072
VERSION AL929072.8 GI:28172212
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
AUTHORS Lovell,J.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk
On Jan 30, 2003 this sequence version replaced gi:28144694.
COMMENT
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfsh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YVC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
beginning 'Dr' were identified by the Recon repeat discovery system
(Zhigong Bao and Sean Eddy, submitted), and those beginning 'dtr'
were identified by Rick Waterman (Stephen Johnson lab, WashU). For
further information see http://www.Projects/D_rerio/fishmark.shtml
CH211-140B8 is from a CHORI-211 BAC library
VECTOR: pTARBAC2.1.
FEATURES
source
1..160432
Location/Qualifiers
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-140B8"
/clone_lib="CHORI-211"
ORIGIN
Query Match 83.2%; Score 20.8; DB 5; Length 160432;
Best Local Similarity 91.7%; Pred. No. 6.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 50
AC124814/c
LOCUS AC124814 162020 bp DNA linear ROD 01-APR-2004
DEFINITION Mus musculus chromosome 1, clone RP24-119P5, complete sequence.
AC124814
VERSION AC124814.10 GI:45917479
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurgachii; Murioidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE Mus musculus chromosome 1, clone RP24-119P5
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 162020)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santoe, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS 3 (bases 1 to 162020)
Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Rachupka, A., Ramessamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubb, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
AUTHORS 4 (bases 1 to 162020)
Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,

Dearellano, K., Diaz, J., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Rachupka, A., Ramessamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubb, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (01-APR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 1, 2004 this sequence version replaced gi:42734526.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L25056
Center clone name: 119_P_5

FEATURES
source
1..162020
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="1"
/map="1"
/clone="RP24-119P5"
/clone_11b="RP24-119P5"
1..334
/note="wgs end extension
/note="wgs end extension
/note="wgs end extension
/note="wgs end extension
/note="clone boundary
clone_end:SP6
site:Mbol
repeat_region
432..455
/rpt_family="GC-rich"
complement(1076..1216)
/rpt_family="B1_MM"
1456..1498
/rpt_family="B1_MM"
repeat_region
1456..1498
/rpt_family="B4"
1499..1629
/rpt_family="B1_MM"
repeat_region
1630..1652
/rpt_family="(CAA)n"
complement(1943..2142)
/rpt_family="B3"
repeat_region
2875..2984
/rpt_family="B1_MM"
complement(3287..3486)
/rpt_family="B3"
3765..3905
/rpt_family="B1_MM"
repeat_region
4557..4591
/rpt_family="(T)n"
repeat_region
5024..5074
/rpt_family="CT-rich"
6078..6116
/rpt_family="(TTC)n"

```
repeat_region      6452. .6473      /rpt_family="GC_rich"
repeat_region      6605. .6654      /rpt_family="TGG)n"
repeat_region      6655. .6721      /rpt_family="(GA)n"
repeat_region      9168. .9228      /rpt_family="B1_MM"
repeat_region      14017. .14045      /rpt_family="GA-rich"
repeat_region      16169. .16196      /rpt_family="AT_rich"
repeat_region      17317. .17359      /rpt_family="(T)n"
repeat_region      20039. .20180      /rpt_family="(CA)n"
repeat_region      20854. .21118      /rpt_family="B1F"
repeat_region      21138. .21221      /rpt_family="(CA)n"
repeat_region      21308. .21407      /rpt_family="(CA)n"
repeat_region      22195. .22233      /rpt_family="Charlie9"
repeat_region      22233. .22291      /rpt_family="(TC)n"
repeat_region      23356. .23584      /rpt_family="(TGG)n"
repeat_region      22634. .22674      /rpt_family="B4A"
repeat_region      23556. .23584      /rpt_family="(CAGA)n"
repeat_region      24120. .24162      /rpt_family="(CGTG)n"
repeat_region      24964. .25069      /rpt_family="B3A"
repeat_region      25295. .25334      /rpt_family="(CTG)n"
repeat_region      25369. .25427      /rpt_family="PB1D9"
repeat_region      25954. .25954      /rpt_family="ID B1"
repeat_region      26507. .26507      /rpt_family="UR1A"
repeat_region      28010. .28094      /rpt_family="(GA)n"
repeat_region      28912. .28912      /rpt_family="UR1A"
repeat_region      29110. .29171      /rpt_family="(TCCC)n"
```

```
Query Match      83.2%: Score 20.8; DB 9; Length 162020;
Best Local Similarity 91.7%: Pred. No. 6.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      1 AAAAAAAAACTCATCTTTAA 24
          |||||
Db      47796 AAAAAAAAACTCATCTTTAA 47773
```

Search completed: December 14, 2005, 11:11:59
Job time : 875.8 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 ; Search time 203.2 Seconds
(without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-9

Perfect score: 25
Sequence: 1 aaaaaaaaaaccctcatcattcaaac 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 150 summaries

Database :

N_Geneseq_21: *
1: Geneseq1980s: *
2: Geneseq1990s: *
3: Geneseq2000s: *
4: Geneseq2001as: *
5: Geneseq2001bs: *
6: Geneseq2002as: *
7: Geneseq2002bs: *
8: Geneseq2003as: *
9: Geneseq2003bs: *
10: Geneseq2003cs: *
11: Geneseq2003ds: *
12: Geneseq2004as: *
13: Geneseq2004bs: *
14: Geneseq2005s: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.8	87.2	480	6	ABK45943
2	20.8	83.2	481	9	ACH23934
3	20.8	83.2	5203	6	ABL32919
4	20.8	83.2	37487	9	ADA02498
5	20.8	83.2	37487	10	ADB72236
6	20.8	83.2	37487	10	ADBE5748
7	20.8	83.2	37487	10	ADBE5748
8	20.4	81.6	121001	3	AAE22284
9	20.2	80.8	2754	3	AAA63944
10	20.2	80.8	5317	6	ABL32609
11	20.2	80.8	5430	4	AA546291
12	20.2	80.8	6880	6	ABL31322
13	20.2	80.8	6880	6	ABL70293
14	20.2	80.8	6880	6	AA561223
15	20.2	80.8	9369	6	ABL17980
16	20.2	80.8	10480	6	ABL34200
17	20.2	80.8	12366	6	ABL03578
18	20.2	80.8	57860	12	ADG97692
19	20.2	80.8	110000	14	ABE39175_12

20	20.2	80.8	110000	14	ABE42401_12	Continuation (13 o
21	20.2	80.8	184368	14	ABE35722	Ab35722 L. pneumo
22	20	80.0	3679	10	ADD45743	Add45743 Human gen
23	20	80.0	3679	10	ADE59936	Ades59936 Human gen
24	19.8	79.2	5919	6	ABL33388	AB133388 Human imm
25	19.8	79.2	5919	6	ABK31194	ABK31194 Signal tr
26	19.8	79.2	5919	6	ABL70521	ABL70521 Chemicall
27	19.8	79.2	5919	6	AA561106	AA561106 Human gen
28	19.8	79.2	6181	6	ABK31367	ABK31367 Signal tr
29	19.8	79.2	6181	6	ABL70324	ABL70324 Chemicall
30	19.8	79.2	6181	6	AA561271	AA561271 Human gen
31	19.8	79.2	6676	6	ABQ67014	ABQ67014 Human ang
32	19.8	79.2	10254	6	ABQ67046	ABQ67046 Human ang
33	19.8	79.2	122937	13	ABD33516	ABD33516 Human ang
34	19.8	79.2	160198	10	ADL13962	ADL13962 Oryzomorph
35	19.4	77.6	2000	6	ABZ17257	ABZ17257 Arabidops
36	19.4	77.6	68295	14	ABE76936	ABE76936 Chicken o
37	19.4	77.6	75815	14	ADM44335	ADM44335 Chicken o
38	19.4	77.6	75815	14	AEA00969	AEA00969 DNA inser
39	19.4	77.6	75815	14	ABE76922	ABE76922 Chicken o
40	19.4	77.6	77872	14	ADM44343	ADM44343 Oryzomorph
41	19.4	77.6	77872	14	AEA00977	AEA00977 Oryzomorph
42	19.4	77.6	77872	14	ABE76930	ABE76930 Chicken o
43	19.2	76.8	151	4	AAH37044	AAH37044 Human col
44	19.2	76.8	247	8	ABX54379	ABX54379 Bovine ES
45	19.2	76.8	500	13	ACF85962	ACF85962 Human SIR
46	19.2	76.8	541	4	ABA62042	ABA62042 Human foe
47	19.2	76.8	541	4	AA141981	AA141981 Probe #10
48	19.2	76.8	584	6	ABQ39865	ABQ39865 Oligonuc
49	19.2	76.8	584	6	ABQ39864	ABQ39864 Oligonuc
50	19.2	76.8	1620	3	AAA98852	AAA98852 Human dif
51	19.2	76.8	1620	3	AAA98133	AAA98133 Human dif
52	19.2	76.8	1620	6	AAD29963	AAD29963 Human pha
53	19.2	76.8	2113	4	AAK94674	AAK94674 Human ful
54	19.2	76.8	2113	12	ADL31643	ADL31643 Pull leng
55	19.2	76.8	2501	8	ABZ10183	ABZ10183 Haematopo
56	19.2	76.8	2501	8	ABZ10037	ABZ10037 Haematopo
57	19.2	76.8	2767	10	ADG42381	ADG42381 DNA encod
58	19.2	76.8	2898	13	AD547429	AD547429 Bacterial
59	19.2	76.8	4501	14	ADN24544	ADN24544 TIMP3 gen
60	19.2	76.8	4501	14	ADN24550	ADN24550 TIMP3 gen
61	19.2	76.8	4501	14	AEA18045	AEA18045 Converted
62	19.2	76.8	4501	14	AEA18065	AEA18065 Converted
63	19.2	76.8	5376	6	ABL34151	ABL34151 Human imm
64	19.2	76.8	5447	4	AA546758	AA546758 Tumour su
65	19.2	76.8	5474	6	ABL33795	ABL33795 Human imm
66	19.2	76.8	5632	6	ABN80187	ABN80187 Human che
67	19.2	76.8	5926	6	ABL33652	ABL33652 Human imm
68	19.2	76.8	5982	6	AA561403	AA561403 Human gen
69	19.2	76.8	5982	6	ABL32180	ABL32180 Human imm
70	19.2	76.8	6030	6	AA545456	AA545456 Chemicall
71	19.2	76.8	6030	6	ABK28312	ABK28312 DNA trans
72	19.2	76.8	6030	6	AA561348	AA561348 Human gen
73	19.2	76.8	6031	4	AA546622	AA546622 Tumour su
74	19.2	76.8	6074	6	ABK28373	ABK28373 DNA trans
75	19.2	76.8	6099	6	ABK31433	ABK31433 Signal tr
76	19.2	76.8	6099	6	ABL70398	ABL70398 Chemicall
77	19.2	76.8	6099	6	AA561346	AA561346 Human gen
78	19.2	76.8	6200	4	AA564641	AA564641 Human gen
79	19.2	76.8	6335	11	ADW21706	ADW21706 Rat hepat
80	19.2	76.8	6335	6	ABL32588	ABL32588 Human imm
81	19.2	76.8	6634	6	ABL44034	ABL44034 Human imm
82	19.2	76.8	6664	10	ADBS4322	ADBS4322 Pretreat
83	19.2	76.8	7057	4	AA546602	AA546602 Tumour su
84	19.2	76.8	7057	6	ABL33849	ABL33849 Human imm
85	19.2	76.8	7057	6	ABL4595	ABL4595 Human met
86	19.2	76.8	7057	6	ABL70408	ABL70408 Chemicall
87	19.2	76.8	7057	6	AA561355	AA561355 Human gen
88	19.2	76.8	7057	7	AD599856	AD599856 Complem
89	19.2	76.8	7309	4	AA546568	AA546568 Tumour su
90	19.2	76.8	7309	6	ABL33817	ABL33817 Human imm
91	19.2	76.8	7309	6	ABK34011	ABK34011 Human DNA
92	19.2	76.8	7309	10	ADBS4292	ADBS4292 Pretreat

PR 30-JUN-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
DR
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX
PS Claim 1; SEQ ID NO 1146; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=20030073623
XX
SQ Sequence 481 BP; 154 A; 80 C; 67 G; 172 T; 0 U; 8 Other;
Query Match 83.2%; Score 20.8; DB 9; Length 481;
Best Local Similarity 91.7%; Pred. No. 2.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 458 AAAAAAAAAACCTTCATCATTTAA 435
RESULT 3
ABL32919/c
ID ABL32919 standard; DNA; 5203 BP.
XX
XX ABL32919;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Human immune system associated gene SEQ ID NO: 892.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosstatic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
XX Homo sapiens.
OS
XX
XX WO200200928-A2.
PN
XX
XX 03-JAN-2002.
PD
XX

PF 02-JUN-2001; 2001WO-EP007537.
XX
XX
XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
XX
PS Claim 1; SEQ ID NO 892; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5203 BP; 1380 A; 121 C; 1148 G; 2554 T; 0 U; 0 Other;
Query Match 83.2%; Score 20.8; DB 6; Length 5203;
Best Local Similarity 91.7%; Pred. No. 2.6e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 2966 AAAAAAAAAACCTTCATCATTTAAAC 2943
RESULT 4
ADA02498/c
ID ADA02498 standard; DNA; 37487 BP.
XX
XX ADA02498;
AC
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Human MYC carcinoma associated gene, SEQ ID NO:1016.
DE
XX
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytosstatic; gene therapy; drug screening;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO2003057146-A2.
PN
XX
XX 17-JUL-2003.
PD
XX
XX 26-DEC-2002; 2002WO-US041414.
PP
XX
XX 26-DEC-2001; 2001US-00035832.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
XX Morris DW;
PI
XX
XX WPI; 2003-587068/55.
DR
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
PT
XX
XX Claim 1; SEQ ID NO 1016; 245pp; English.
PS
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic

CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a bioclip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 37487 BP; 11370 A; 7915 C; 8285 G; 9917 T; 0 U; 0 Other;
SQ

Query Match 83.2%; Score 20.8; DB 9; Length 37487;
Best Local Similarity 91.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 AAAAAAAAACTTCATCATTTAAA 24
Db 36890 AAAAAACATACCTTCATCATTTAAA 36867

RESULT 5
ADB72236/c
ID ADB72236 standard; DNA; 37487 BP.
XX
AC ADB72236;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human MYC gene.
XX
KW human; d; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Homo sapiens.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
PR 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-239337/23.
XX
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS Claim 1; SEQ ID NO 64; 2304pp; English.
XX
CC The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined

CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX

Sequence 37487 BP; 11370 A; 7915 C; 8285 G; 9917 T; 0 U; 0 Other;
SQ

Query Match 83.2%; Score 20.8; DB 10; Length 37487;
Best Local Similarity 91.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 AAAAAAAAACTTCATCATTTAAA 24
Db 36890 AAAAAACATACCTTCATCATTTAAA 36867

RESULT 6
ADE82938/c
ID ADE82938 standard; DNA; 37487 BP.
XX
AC ADE82938;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human MYC genomic DNA sequence.
XX
KW human; cancer-associated nucleic acid; screening; cancer; lymphoma;
KW leukemia; breast cancer; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO2003080808-A2.
XX
PD 02-OCT-2003.
XX
PF 21-MAR-2003; 2003WO-US008919.
XX
PR 21-MAR-2002; 2002US-0367025P.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-865119/80.
XX
PT New cancer-associated proteins and nucleic acids, useful for screening
PT for anticancer activity in a potential drug, or for detecting,
PT diagnosing, preventing and treating cancers, e.g. lymphoma, leukemia or
PT breast cancer.
XX
PS Claim 16; SEQ ID NO 22; 248pp; English.
XX
CC The invention comprises human and mouse cancer-associated nucleic acid
CC sequences. The cancer associated nucleic acids of the invention are
CC useful for screening for anticancer activity in a potential drug, as well
CC as detecting, diagnosing, preventing and treating cancers (e.g. lymphoma,
CC leukemia, or breast cancer). The present sequence represents a cancer-
CC associated nucleic acid of the invention.
XX

Sequence 37487 BP; 11370 A; 7914 C; 8286 G; 9917 T; 0 U; 0 Other;
SQ

Query Match 83.2%; Score 20.8; DB 10; Length 37487;
Best Local Similarity 91.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 1 AAAAAAAAACTTCATCATTTAAA 24
Db 36890 AAAAAACATACCTTCATCATTTAAA 36867

RESULT 7
ADE95746/c


```
ID ADE95746 standard; DNA; 37487 BP.
XX
XX ADE95746;
XX
XX 12-FEB-2004 (first entry)
XX
DE Human MYC gene genomic DNA sequence.
XX
XX cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
XX lymphoma; breast cancer; prostate cancer; leukemia; ds; human; MYC.
XX
XX Homo sapiens.
XX
XX MO2003039484-A2.
XX
XX 15-MAY-2003.
XX
XX 08-NOV-2002; 2002MO-US036071.
XX
XX 08-NOV-2001; 2001US-00052482.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW, Engelhard EK;
XX
XX WPI; 2003-441462/41.
XX
XX New carcinoma associated nucleic acids and proteins, useful for screening
XX drug candidates, or for diagnosing and treating carcinomas, e.g.
XX lymphoma, breast cancer, prostate cancer or leukemia.
XX
XX Claim 1; SEQ ID NO 4; 793bp; English.
XX
XX This invention relates to novel recombinant nucleic acids for use in
XX diagnosis and treatment of cancer, especially carcinomas, as well as the
XX use of compositions in screening methods. The compositions of the
XX invention may have cytostatic activity whilst the disclosed sequences may
XX be useful for gene therapy. The carcinoma associated nucleic acids and
XX proteins are useful for diagnosing and treating carcinomas, for example
XX lymphoma, breast cancer, prostate cancer or leukemia, or for screening
XX drug candidates or bioactive agents capable of binding to, or modulating
XX the activity of, a carcinoma associated protein. The present sequence is
XX the genomic DNA sequence of the human MYC gene which is a carcinoma
XX associated gene of the invention.
XX
XX Sequence 37487 BP; 11370 A; 7915 C; 8285 G; 9917 T; 0 U; 0 Other;
XX
XX Query Match 83.2%; Score 20.8; DB 10; Length 37487;
XX Best Local Similarity 91.7%; Pred. No. 2.7e+02;
XX Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAACTTCATCATTTAA 24
XX ||||| ||||| ||||| |||||
XX ||||| ||||| ||||| |||||
XX 36890 AAAAAACATCACTTCATCATTTAA 36867
XX
XX RESULT 8
XX ID AAF22284 standard; DNA; 121001 BP.
XX
XX AAF22284;
XX
XX 20-MAR-2001 (first entry)
XX
XX BAC containing repeats from centromeres 1-4 #7.
XX
XX Centromere; microsome; vector; ds.
XX
XX Arabidopsis thaliana.
XX
XX MO200055325-A2.
XX
XX 21-SEP-2000.
```

```
XX
XX 17-MAR-2000; 2000MO-US007392.
XX
XX 18-MAR-1999; 99US-0125219P.
XX
XX 01-APR-1999; 99US-0127409P.
XX
XX 18-MAY-1999; 99US-0134770P.
XX
XX 13-SEP-1999; 99US-0135844P.
XX
XX 17-SEP-1999; 99US-0154603P.
XX
XX 16-DEC-1999; 99US-0172493P.
XX
XX (UYCH-) UNIV CHICAGO.
XX
XX Preuss D, Copenhaver G, Keith K;
XX
XX WPI; 2000-587529/55.
XX
XX Recombinant DNA construct comprising a plant centromere, useful for
XX producing stably inherited microsome which can serve as vectors for the
XX construction of transgenic plant and animal cells.
XX
XX Claim 102; Page 404-431; 1449bp; English.
XX
XX The present invention relates to a recombinant DNA construct of a plant
XX (Arabidopsis thaliana) centromere. The constructs are useful for
XX producing stably inherited microsome which can serve as vectors for the
XX construction of transgenic plant and animal cells expressing selected
XX proteins such as hormones, enzymes, interleukins, clotting factors,
XX cytokines, antibodies, and growth factors
XX
XX Sequence 121001 BP; 39743 A; 24748 C; 22105 G; 33637 T; 0 U; 768 Other;
XX
XX Query Match 81.6%; Score 20.4; DB 3; Length 121001;
XX Best Local Similarity 95.5%; Pred. No. 3.9e+02;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAACTTCATCATTTA 22
XX ||||| ||||| ||||| |||||
XX ||||| ||||| ||||| |||||
XX 115019 AAAAAAAAACTTCATCATTTCA 115040
XX
XX RESULT 9
XX ID AAA63944 standard; cDNA; 2754 BP.
XX
XX AAA63944;
XX
XX 04-DEC-2000 (first entry)
XX
XX cDNA encoding a pancreatic ductal trophic factor (PDTF).
XX
XX Pancreatic ductal trophic factor; PDTF; mesenchyme cell;
XX transforming growth factor-beta; TGF-beta; pancreatic tissue growth;
XX pancreatic duct tissue outgrowth; diabetes; ss.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX CDS 504..1652
XX FT /*tag= a
XX FT /product= "a pancreatic ductal trophic factor (PDTF)"
XX FT sig_peptide 504..566
XX FT /*tag= b
XX FT mat_peptide 567..1649
XX FT /*tag= c
XX
XX MO200047243-A1.
XX
XX 17-AUG-2000.
XX
XX 10-FEB-2000; 2000MO-US003424.
XX
XX 10-FEB-1999; 99US-0119574P.
XX
```

PA (ONTO-) ONTOGENY INC.
 XX
 PI Guicherit O, Pang K, Wang M;
 XX
 DR WPI: 2000-524486/47.
 DR P-PSDB; AAB08338.
 XX
 PT Inducing outgrowths of pancreatic duct tissue, using pancreatic ductal
 PT trophic factors expressed by e14.5 pancreatic mesenchyme cells, for the
 PT treatment of diabetes.
 XX
 PS Disclosure; Page 73-75; 76pp; English.
 XX
 CC The present sequence encodes a pancreatic ductal trophic factor (PDTF)
 CC polypeptide. The polypeptide is expressed by e14.5 pancreatic mesenchyme
 CC cells (especially members of the TGF-beta (transforming growth factor-
 CC beta) super family), and is capable of stimulating growth and
 CC differentiation of pancreatic tissue. The polypeptides may be used for
 CC inducing outgrowths of pancreatic duct tissue, e.g. for treating diabetes
 XX
 SQ Sequence 2754 BP; 905 A; 525 C; 599 G; 725 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 3; Length 2754;
 Best Local Similarity 88.0%; Pred. No. 4.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTTCATCATTTAAAC 25
 DB 2664 AAAAAAAAACTTCATCATTTAAAC 2640
 RESULT 10
 ABL32609/c
 ID ABL32609 standard; DNA; 5317 BP.
 XX
 AC ABL32609;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 582.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cyostatic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.
 KW
 XX
 OS Homo sapiens.
 XX
 PI WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 582; 32pp + Sequence Listing; German.
 XX

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 5317 BP; 1085 A; 225 C; 1468 G; 2539 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 6; Length 5317;
 Best Local Similarity 88.0%; Pred. No. 4.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTTCATCATTTAAAC 25
 DB 2161 AAAAAAAAACTTCATCATTTAAAC 2137
 RESULT 11
 AAS46291/c
 ID AAS46291 standard; DNA; 5430 BP.
 XX
 AC AAS46291;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #13.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cyostatic; cancer;
 KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PI WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP002955.
 XX
 PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumor suppressor
 PT genes and oncogenes, useful in designing primers and probes for analyzing
 PT diseases associated with cytosine methylation state e.g. cancer.
 XX
 PS Claim 1; SEQ ID NO 13; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulfite, of genes associated with tumour suppression and oncogenes
 CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
 CC 500 are missing from the sequence listing) sequences (58) and sequences
 CC complementary to (58). The nucleic acid may be a peptide nucleic acid
 CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
 CC probes for detecting the cytosine methylation state and/or single
 CC nucleotide polymorphisms and also to be used in an array for analysing
 CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
 CC probes can also be used in a method for ascertaining genetic and/or
 CC epigenetic parameters for the diagnosis and/or therapy of existing
 CC diseases or the predisposition to specific diseases, by analysing

CC cytosine methylations. The parameters may be compared to another set of
 CC genetic and/or epigenetic parameters, the differences serving as basis
 CC for diagnosis and/or prognosis events which are disadvantageous to
 CC patients. The present sequence is one of the 533 genomic sequences
 CC derived from tumour suppressor genes and oncogenes. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 5430 BP; 1740 A; 36 C; 1090 G; 2564 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 4; Length 5430;
 Best Local Similarity 88.0%; Pred. No. 4.3e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTTCATCATTTAAAC 25
 Db 4655 AAAAAAAAACTTCATCATTTAAAC 4631
 RESULT 12
 ID ABL70293 standard; DNA; 6880 BP.
 AC ABL70293;
 DT 23-APR-2002 (first entry)
 DE Signal transduction associated gene modified DNA #83.
 DE
 KW Human; signal transduction associated gene; cytosine methylation state;
 KW Cpg island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 OS
 OS Homo sapiens.
 OS Synthetic.
 FN WO200200926-A2.
 PD 03-JAN-2002.
 PE 29-JUN-2001; 2001WO-EP007472.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 WP1; 2002-147896/19.
 PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction.
 PS Claim 1; SEQ ID NO 165; 24pp; English.
 CC The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
 CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
 CC the cytosine methylation state (Cpg islands) of these genes, and a method
 CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
 CC genes associated with signal transduction. The genomic DNA can be
 CC obtained from cells or cellular components which contain DNA, e.g. cell
 CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
 CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
 CC brain, heart, prostate, lung, breast or liver, histologic object slides,
 CC and all their possible combinations. The sequences of the invention are
 CC useful for the diagnosis and therapy of diseases associated with signal
 CC transduction e.g. solid tumours and cancer. ABL70293-ABL70294 represent
 CC chemically pretreated genomic DNA sequences of different genes associated

CC with signal transduction, or their complementary sequences. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office
 CC
 SQ Sequence 6880 BP; 2032 A; 49 C; 1380 G; 3419 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 6; Length 6880;
 Best Local Similarity 88.0%; Pred. No. 4.4e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAACTTCATCATTTAAAC 25
 Db 796 AAAAAAAAACTTCATCATTTATAC 772
 RESULT 13
 ID ABL70293 standard; DNA; 6880 BP.
 AC ABL70293;
 DT 01-JUN-2002 (first entry)
 DE Chemically treated cell signalling DNA sequence#92.
 DE
 KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
 KW tumour; cytostatic; ds.
 OS
 OS Unidentified.
 OS
 FN WO200202807-A2.
 PD 10-JAN-2002.
 PE 29-JUN-2001; 2001WO-EP007471.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 WP1; 2002-154758/20.
 PT Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signalling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signalling.
 PS Claim 1; SEQ ID NO 183; 24pp + Sequence Listing; English.
 CC The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signalling. The activity of the modified sequences of the
 CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signalling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signalling. Note: The
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office
 CC
 SQ Sequence 6880 BP; 2032 A; 49 C; 1380 G; 3419 T; 0 U; 0 Other;
 Query Match 80.8%; Score 20.2; DB 6; Length 6880;
 Best Local Similarity 88.0%; Pred. No. 4.4e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTTCATTTAAAC 25
Db 796 AAAAAAAAACTTCATTTAAAC 772

RESULT 14
AAS61223/c
ID AAS61223 standard; DNA; 6880 BP.

AC AAS61223;

DT 29-JAN-2002 (first entry)

DE Human gene regulation-associated gene oligonucleotide #178.

XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumor;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotoxic; gynecological; anti-tumour; immunosuppressive; cytostatic.

XX Homo sapiens.

OS WO20017375-A2.

PN 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP003968.

PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIDEMIOLOGICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-017470/02.

XX New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
and therapy of diseases e.g. severe combined immunodeficiency disease.

XX Claim 1; SEQ ID NO 183; 26pp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 6880 BP; 2032 A; 49 C; 1380 G; 3419 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 6; Length 6880;

Best Local Similarity 88.0%; Pred. No. 4.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTTCATTTAAAC 25
Db 796 AAAAAAAAACTTCATTTAAAC 772

RESULT 15

ABL17980
ID ABL17980 standard; DNA; 9369 BP.

AC ABL17980;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 5413.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter UC, Adams M, Li PWD, Myers BW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

XX Claim 1; SEQ ID NO 5413; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (AB557737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 9369 BP; 2868 A; 2217 C; 1951 G; 2333 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 4; Length 9369;

Best Local Similarity 88.0%; Pred. No. 4.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATTTAAAC 25
Db 321 AAAAAAAAACTTCATTTAAAC 345

RESULT 16

ABL34200/c
ID ABL34200 standard; DNA; 10480 BP.

XX ABL34200;

DT 26-MAR-2002 (first entry)

```

XX DE Human immune system associated gene SEQ ID NO: 2173.
XX
XX KM Human, immune system disease; cytosine methylation; antiasthmatic;
XX KM antiarteriosclerotic; antianaemic; cytosolic; noctropic;
XX KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KM antirheumatic; antirheumatic; antidiabetic; antiparasitic;
XX KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200200928-A2.
XX
XX PD 03-JAN-2002.
XX
XX PF 02-JUL-2001; 2001WO-EP007537.
XX
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX
XX PA (EPIC-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX
XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
XX PT diagnosis and treatment of diseases associated with abnormal cytosine
XX PT methylation.
XX
XX PS Claim 1; SEQ ID NO 2173; 32pp + Sequence Listing; German.
XX
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention
XX
XX SQ Sequence 10480 BP; 2426 A; 182 C; 2445 G; 5427 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 6; Length 10480;
XX Best Local Similarity 88.0%; Pred. No. 4.4e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAACTTCATCATTTAAAC 25
XX DB 7716 AAAAAAAAACTTCATCATTTAAAC 7692
XX
XX RESULT 17
XX ABL03578
XX ID ABL03578 standard; CDNA; 12366 BP.
XX
XX AC ABL03578;
XX
XX XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5216.
XX
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical; gene; ss.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX

```

```

XX PF 23-MAR-2001; 2001WO-US009231.
XX
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB59475.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX
XX PS Claim 1; SEQ ID NO 5216; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 12366 BP; 3696 A; 2959 C; 2662 G; 3049 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 4; Length 12366;
XX Best Local Similarity 88.0%; Pred. No. 4.4e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAACTTCATCATTTAAAC 25
XX DB 321 AAAAAAAAAAGATTCACATTTAAAC 345
XX
XX RESULT 18
XX ADQ97692/C
XX ID ADQ97692 standard; DNA; 57860 BP.
XX
XX AC ADQ97692;
XX
XX DT 07-OCT-2004 (first entry)
XX
XX DE Mouse cancer associated sequence MD10-029, SEQ ID 669.
XX
XX KM Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
XX OS Mus musculus.
XX
XX PN WO2004060304-A2.
XX PD 22-JUL-2004.
XX
XX PF 22-DEC-2003; 2003WO-US041389.
XX PR 27-DEC-2002; 2002US-00330773.
XX PA (SAGR-) SAGRES DISCOVERY INC.
XX
XX PI Morris DW, Malandro MS;
XX DR WPI; 2004-543781/52.
XX
XX PT New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX

```

PS Claim 1; SEQ ID NO 669; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (AD097025-
CC AD098004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 57860 BP; 16311 A; 11770 C; 11886 G; 17541 T; 0 U; 352 Other;

Query Match 80.8%; Score 20.2; DB 12; Length 57860;
Best Local Similarity 88.0%; Pred. No. 4.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTAAC 25
Db 15110 AAAAAAAAACTTCATCTTAAC 15086

RESULT 19
AEB39175_12
Continuation (13 of 35) of AEB39175 from base 1200001 (L. pneumophila DNA SEQ ID NO 3507
WP Sequence split into 35 fragments LOCUS AEB39175 Accession Aeb39175
WP Fragment Name Begin End
WP AEB39175_00 1 110000
WP AEB39175_01 100001 210000
WP AEB39175_02 200001 310000
WP AEB39175_03 300001 410000
WP AEB39175_04 400001 510000
WP AEB39175_05 500001 610000
WP AEB39175_06 600001 710000
WP AEB39175_07 700001 810000
WP AEB39175_08 800001 910000
WP AEB39175_09 900001 1010000
WP AEB39175_10 1000001 1110000
WP AEB39175_11 1100001 1210000
WP AEB39175_12 1200001 1310000
WP AEB39175_13 1300001 1410000
WP AEB39175_14 1400001 1510000
WP AEB39175_15 1500001 1610000
WP AEB39175_16 1600001 1710000
WP AEB39175_17 1700001 1810000
WP AEB39175_18 1800001 1910000
WP AEB39175_19 1900001 2010000
WP AEB39175_20 2000001 2110000
WP AEB39175_21 2100001 2210000
WP AEB39175_22 2200001 2310000
WP AEB39175_23 2300001 2410000
WP AEB39175_24 2400001 2510000
WP AEB39175_25 2500001 2610000
WP AEB39175_26 2600001 2710000
WP AEB39175_27 2700001 2810000
WP AEB39175_28 2800001 2910000
WP AEB39175_29 2900001 3010000
WP AEB39175_30 3000001 3110000
WP AEB39175_31 3100001 3210000
WP AEB39175_32 3200001 3310000
WP AEB39175_33 3300001 3410000
WP AEB39175_34 3400001 3503610

Query Match 80.8%; Score 20.2; DB 14; Length 110000;
Best Local Similarity 88.0%; Pred. No. 4.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTAAC 25
Db 60303 AAGAACCAACCATCATCTTAAC 60327

RESULT 20
AEB42401_12
Continuation (13 of 34) of AEB42401 from base 1200001 (L. pneumophila DNA SEQ ID NO 6733

WP Sequence split into 34 fragments LOCUS AEB42401 Accession Aeb42401
WP Fragment Name Begin End
WP AEB42401_00 1 110000
WP AEB42401_01 100001 210000
WP AEB42401_02 200001 310000
WP AEB42401_03 300001 410000
WP AEB42401_04 400001 510000
WP AEB42401_05 500001 610000
WP AEB42401_06 600001 710000
WP AEB42401_07 700001 810000
WP AEB42401_08 800001 910000
WP AEB42401_09 900001 1010000
WP AEB42401_10 1000001 1110000
WP AEB42401_11 1100001 1210000
WP AEB42401_12 1200001 1310000
WP AEB42401_13 1300001 1410000
WP AEB42401_14 1400001 1510000
WP AEB42401_15 1500001 1610000
WP AEB42401_16 1600001 1710000
WP AEB42401_17 1700001 1810000
WP AEB42401_18 1800001 1910000
WP AEB42401_19 1900001 2010000
WP AEB42401_20 2000001 2110000
WP AEB42401_21 2100001 2210000
WP AEB42401_22 2200001 2310000
WP AEB42401_23 2300001 2410000
WP AEB42401_24 2400001 2510000
WP AEB42401_25 2500001 2610000
WP AEB42401_26 2600001 2710000
WP AEB42401_27 2700001 2810000
WP AEB42401_28 2800001 2910000
WP AEB42401_29 2900001 3010000
WP AEB42401_30 3000001 3110000
WP AEB42401_31 3100001 3210000
WP AEB42401_32 3200001 3310000
WP AEB42401_33 3300001 3345687

Query Match 80.8%; Score 20.2; DB 14; Length 110000;
Best Local Similarity 88.0%; Pred. No. 4.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTAAC 25
Db 68091 AAGAACCAACCATCATCTTAAC 68115

RESULT 21
AEB35722
ID AEB35722 standard; DNA; 184368 BP.
XX
AC AEB35722;
XX
DT 08-SEP-2005 (first entry)
XX
DE L. pneumophila DNA SEQ ID NO 54.
XX
KW detection; infection; Antibacterial; Vaccine; ds; gene.
XX
OS Legionella pneumophila.
XX
PN W02005049642-A2.
XX
XX 02-JUN-2005.
XX
PD 23-SEP-2004; 2004MO-IB003578.
XX
PF 21-NOV-2003; 2003FR-00013687.
XX
PR
XX
XX (INSP) INST PASTEUR.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX (UPLY-) UNIV LYON 1 BERNARD CLAUDE.
XX (CNRS) CNRS CENT NAT RECH SCI.

PI Buchrieser C, Ticht M, Etienne J, Ma L, Cazale C, Glaser P;
PI Kuenick C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;
PI Jarrard S;
XX
XX WPI; 2005-368305/40.
XX
PT New genome of *Legionella pneumophila* Paris strain and derived
PT polypeptides, useful for detection or identification of the strain and
PT for treatment and prevention of infections.
XX
PS Claim 1; SEQ ID NO 54; 660bp; English.
XX
CC The invention relates to an isolated or purified nucleotide sequences (I)
CC from *Legionella pneumophila* Paris strain. (I), and their related
CC sequences or fragments, are useful as primers and probes for detection
CC and amplification, including differentiation between the Paris and
CC Philadelphia strains of *Legionella pneumophila* and to prepare recombinant
CC (hybrid) polypeptides (II) (II) are also useful for preparation of
CC specific antibodies (Ab), also used for detection/identification of
CC Legionella, and some (I), specifically those involved in synthesis of
CC surface proteins, are targets for identification of inhibitors. (II), or
CC vectors that contain (I), are useful as vaccines and immunogenic
CC compositions, for treatment and prevention of infections by *L.*
CC pneumophila. The present sequence represents a *L. pneumophila* DNA.
XX
SQ Sequence 184368 BP; 57341 A; 33604 C; 36303 G; 57120 T; 0 U; 0 Other;
Query Match 80.8%; Score 20.2; DB 14; Length 184368;
Best Local Similarity 88.0%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 90660 AAGAACAAACATCATCATTTAAAC 90684
RESULT 22
ADD45743
ID ADD45743 standard; DNA; 3679 BP.
XX
AC ADD45743;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene X01703, SEQ ID NO 11411.
XX
XX Human; de; gene; pain; neuronal tissue; gene therapy;
KM spinal segmental nerve injury; chronic constriction injury; CCI;
KM spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
OS
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
DR GENBANK; X01703.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX

PS Claim 1; Page; 1017p; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3679 BP; 913 A; 848 C; 904 G; 1014 T; 0 U; 0 Other;
Query Match 80.0%; Score 20; DB 10; Length 3679;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCTTCATCATTT 20
Db 611 AAAAAAAAAACCTTCATCATTT 630
RESULT 23
ADE59936
ID ADE59936 standard; DNA; 3679 BP.
XX
AC ADE59936;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human gene X01703, SEQ ID NO 5832.
XX
XX Human; de; gene; pain; neuronal tissue; gene therapy;
KM spinal segmental nerve injury; chronic constriction injury; CCI;
KM spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
OS
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX

DR WPI, 2003-268312/26.
 DR GENBANK; X01703.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1, Page; 1017pp, English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
 CC specification) which encodes one of the polypeptides of the invention
 CC which is differentially expressed during pain. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3679 BP; 913 A; 848 C; 904 G; 1014 T; 0 U; 0 Other;
 Query Match 80.0%; Score 20; DB 10; Length 3679;
 Best Local Similarity 100.0%; Pred. No. 5,1e+02;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTTCATCATT 20
 DB 611 AAAAAAAAACTTCATCATT 630
 RESULT 24
 ABL3388/C
 ID ABL32388 standard; DNA; 5919 BP.
 XX
 AC ABL32388;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 361.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-anemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antiprolactic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX

PF 02-JUL-2001; 2001WO-EP007537.
 XX
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIC-) EPIDENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 361; 32pp + Sequence listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 5919 BP; 1524 A; 130 C; 1378 G; 2887 T; 0 U; 0 Other;
 Query Match 79.2%; Score 19.8; DB 6; Length 5919;
 Best Local Similarity 91.3%; Pred. No. 6,1e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTTCATCATTAA 23
 DB 4059 AAAAAAAAACTTCATCATTAA 4037
 RESULT 25
 ABL31194/C
 ID ABL31194 standard; DNA; 5919 BP.
 XX
 AC ABL31194;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified DNA #19.
 XX
 KW Human; signal transduction associated gene; cytosine methylation state;
 KW Cpg island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200200926-A2.
 XX
 PD 03-JAN-2002.
 XX
 DE 29-JUN-2001; 2001WO-EP007472.
 PF 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIC-) EPIDENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-147896/19.
 XX
 PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction.
 XX

PS Claim 1; SEQ ID NO 37; 24pp; English.

XX

CC The present invention relates to chemically modified DNA sequences of

CC signal transduction associated genes. The DNA sequences are chemically

CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.

CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting

CC the cytosine methylation state (CpG islands) of these genes, and a method

CC for the diagnosis and/or therapy of genetic and epigenetic parameters of

CC genes associated with signal transduction. The genomic DNA can be

CC obtained from cells or cellular components which contain DNA, e.g. cell

CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,

CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,

CC brain, heart, prostate, lung, breast or liver, histologic object slides,

CC and all their possible combinations. The sequences of the invention are

CC useful for the diagnosis and therapy of diseases associated with signal

CC transduction e.g. solid tumours and cancer. ABR31158-ABR31545 represent

CC chemically pretreated genomic DNA sequences of different genes associated

CC with signal transduction, or their complementary sequences. Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office

CC

XX

SQ Sequence 5919 BP; 1524 A; 130 C; 1378 G; 2887 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 5919;

Best Local Similarity 91.3%; Pred. No. 6.1e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCATTTAA 23

Db 4059 AAAAAAATACCTTCTTCATTTAA 4037

RESULT 26

ID ABL70521/c

XX ABL70521 standard; DNA; 5919 BP.

XX

AC ABL70521;

XX

DT 01-JUL-2002 (first entry)

XX

DE Chemically treated cell signalling DNA sequence#206.

XX

KW Cell signalling; cytosine methylation; cell signalling disease; cancer;

KW tumour; cytostatic; ds.

XX

OS Unidentified.

OS

XX

PN WO200202807-A2.

XX

PD 10-JAN-2002.

XX

PF 29-JUN-2001; 2001WO-EP007471.

XX

PR 30-JUN-2000; 2000DE-01032529.

XX

PR 01-SEP-2000; 2000DE-01043826.

XX

PA (EPig-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K;

XX

DR WPI; 2002-154758/20.

XX

PT Nucleic acid, useful for diagnosis and therapy of diseases associated

PT with cell signaling e.g. cancer, comprises chemically modified genomic

PT sequences of genes associated with cell signaling.

XX

PS Claim 1; SEQ ID NO 411; 24pp + Sequence Listing; English.

XX

CC The invention relates to a nucleic acid comprising a sequence of at least

CC 18 bases of a segment of chemically pretreated DNA of genes associated

CC with cell signalling. The activity of the modified sequences of the

CC invention may be described as cytostatic. The object of the invention is

CC to provide the chemically modified DNA of genes associated with cell

CC signalling, as well as oligonucleotides and/or PNA-oligomers for

CC detecting cytosine methylations, as well as a method which is

CC particularly suitable for the diagnosis and/or therapy of genetic and

CC epigenetic parameters of genes associated with cell signalling. The

CC chemically modified DNA provided by the invention is useful for diagnosis

CC and therapy of diseases such as solid tumours and cancer. The sequences

CC given in records ABL70111-ABL70626 represent chemically pre-treated

CC genomic DNA's of genes associated with cell signalling. Note: The

CC sequence data for this patent is not represented in the printed

CC specification, but is based on sequence information supplied by the

CC European Patent Office

XX

SQ Sequence 5919 BP; 1524 A; 130 C; 1378 G; 2887 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 5919;

Best Local Similarity 91.3%; Pred. No. 6.1e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCATTTAA 23

Db 4059 AAAAAAATACCTTCTTCATTTAA 4037

RESULT 27

ID AAS61106/c

XX AAS61106 standard; DNA; 5919 BP.

XX

AC AAS61106;

XX

DT 29-JAN-2002 (first entry)

XX

DE Human gene regulation-associated gene oligonucleotide #61.

XX

KW Human; Gene regulation-associated gene; severe combined immunodeficiency;

KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;

KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;

KW renal disease; Preeclampsia; cardiac allograft vascular disease;

KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;

KW immunostimulant; cardiant; anti-inflammatory; coagulant; antiasthmatic;

KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

XX

OS Homo sapiens.

OS

XX

PN WO200177375-A2.

XX

PD 18-OCT-2001.

XX

PF 06-APR-2001; 2001WO-EP003968.

XX

PR 06-APR-2000; 2000DE-01019058.

XX

PR 07-APR-2000; 2000DE-01019173.

XX

PR 30-JUN-2000; 2000DE-01032529.

XX

PR 01-SEP-2000; 2000DE-01043826.

XX

PA (EPig-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K;

XX

DR WPI; 2002-017470/02.

XX

PT New nucleic acid sequences from chemically modified genes associated with

PT gene regulation, useful for analyzing cytosine methylations for diagnosis

PT and therapy of diseases e.g. severe combined immunodeficiency disease.

XX

PS Claim 1; SEQ ID NO 63; 26pp; English.

XX

CC The invention relates to 224 nucleic acid sequences comprising at least

CC 18 bases of a chemically pretreated gene associated with gene regulation

CC selected from 43 known genes (or complementary sequences). The chemical

CC pretreatment converts cytosine bases unmethylated at the 5-position to

CC uracil or another base with hybridisation behaviour dissimilar to

CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,

CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes; kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pub/published_pct_sequences
CC
SQ Sequence 5919 BP; 1524 A; 130 C; 1378 G; 2887 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 5919;
Best Local Similarity 91.3%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 23
Db 4059 AAAAAAAAACTTCATCATTTAA 4037

RESULT 28
ABK1367/c
ID ABK1367 standard; DNA; 6181 BP.

AC ABK1367;
XX
DT 23-APR-2002 (first entry)

DE Signal transduction associated gene modified complementary DNA #105.

XX Human; signal transduction associated gene; cytosine methylation state;
KM CpG island; signal transduction associated disease; solid tumour; cancer;
KM antitumour; cytosstatic; mutant; ds.

XX Homo sapiens.
OS Synthetic.

PN WO200200926-A2.

PD 03-JAN-2002.

PF 29-JUN-2001; 2001MO-EP007472.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPig-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-147896/19.

PT Oligonucleotide for diagnosis and therapy of diseases associated with
signal transduction e.g. cancer, comprises chemically modified genomic
sequences of genes associated with signal transduction.

XX Claim 1; SEQ ID NO 210; 24pp; English.

XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,

CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK1155-ABK1345 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 6181 BP; 1619 A; 163 C; 1277 G; 3122 T; 0 U; 0 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 6181;
Best Local Similarity 91.3%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 23
Db 1904 AAAAAAAAACTTCATCAATTAA 1882

RESULT 29
ABL70324/c
ID ABL70324 standard; DNA; 6181 BP.

XX ABL70324;

XX
DT 01-JUL-2002 (first entry)

DE Chemically treated cell signalling DNA sequence complementary to #107.

XX Cell signalling; cytosine methylation; cell signalling disease; cancer;
KM tumour; cytosstatic; ds.

XX Unidentified.

PN WO200202807-A2.

PD 10-JAN-2002.

PF 29-JUN-2001; 2001MO-EP007471.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPig-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-154758/20.

PT Nucleic acid, useful for diagnosis and therapy of diseases associated
with cell signalling e.g. cancer, comprises chemically modified genomic
sequences of genes associated with cell signalling.

XX Claim 1; SEQ ID NO 214; 24pp + Sequence Listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytosstatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL7011-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office

XX Sequence 6181 BP; 1619 A; 163 C; 1277 G; 3122 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 6; Length 6181;
Best Local Similarity 91.3%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTTCATCATTTAA 23
Db 1904 AAAAAAAAACTTCATCATTTAA 1882

RESULT 30
AAS61271/c
ID AAS61271 standard; DNA; 6181 BP.
XX
XX AAS61271;
AC
XX
XX 29-JAN-2002 (first entry)
DT
XX
XX Human gene regulation-associated gene oligonucleotide #226.
DE
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiac; anti-inflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
XX Homo sapiens.
OS
XX
XX WO200177375-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 06-APR-2001; 2001WO-EP003968.
PF
XX
XX 06-APR-2000; 2000DE-01019058.
PR
XX 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
PR
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2002-017470/02.
DR
XX
XX New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.
XX
XX
XX Disclosure; SEQ ID NO 232; 26bp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmetlylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 6181 BP; 1619 A; 163 C; 1277 G; 3122 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 6; Length 6181;
Best Local Similarity 91.3%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTTCATCATTTAA 23
Db 1904 AAAAAAAAACTTCATCATTTAA 1882

RESULT 31
ABO67014/c
ID ABO67014 standard; DNA; 6676 BP.
XX
XX ABO67014;
AC
XX
XX 28-AUG-2002 (first entry)
DT
XX
XX Human angiogenesis associated polynucleotide SEQ ID NO 44.
DE
XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
XX
XX Homo sapiens.
OS
XX
XX WO200246454-A2.
PN
XX
XX 13-JUN-2002.
PD
XX
XX 06-DEC-2001; 2001WO-EP014320.
PF
XX
XX 06-DEC-2000; 2000DE-01061338.
PR
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
XX Schacht O;
PI
XX
XX WPI; 2002-500450/53.
DR
XX
XX New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
XX
XX Claim 1; SEQ ID NO 44; 41bp + Sequence Listing; German.

XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABO66971-ABO67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX
XX Sequence 6676 BP; 1891 A; 55 C; 1377 G; 3353 T; 0 U; 0 Other;
SQ
Query Match 79.2%; Score 19.8; DB 6; Length 6676;
Best Local Similarity 91.3%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAAAAAACTTCATCATTTAA 24
Db 5972 AAAAAAAAACTTCATCATTTAA 5950

RESULT 32
ABQ67046/c
ID ABQ67046 standard; DNA; 10254 BP.
XX
AC ABQ67046;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 76.
XX
KM Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KM inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
KM macular degeneration; inflammatory bowel disease; Crohn's disease;
KM antipneumatic; antiarthritic; antidiabetic; antipsoriatic;
KM antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 76; 41pp + Sequence Listing; German.
XX
XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC prioritis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 10254 BP; 2837 A; 106 C; 2239 G; 5072 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 6; Length 10254;
Best Local Similarity 91.3%; Pred. No. 6.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCTGATCATTTAA 23
DB 10104 AAAAAAAAAACTGATCATTTAA 10082
XX
RESULT 33
ABD33516
ID ABD33516 standard; DNA; 122937 BP.
XX
AC ABD33516;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-101.

XX
KM Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KM ds; cancer; cytosstatic.
XX
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 694; 182pp; English.
XX
XX The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 122937 BP; 38411 A; 20858 C; 22659 G; 41009 T; 0 U; 0 Other;
XX
Query Match 79.2%; Score 19.8; DB 13; Length 122937;
Best Local Similarity 91.3%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAAAAAAAAACCTGATCATTTAA 24
DB 23089 AAAAAAAAACTGATCATTTAA 29111
XX
RESULT 34
ADL13962/c
ID ADL13962 standard; DNA; 160198 BP.
XX
AC ADL13962;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #494.
XX
KM ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KM joint space narrowing; osteophyte development; joint pain;
KM osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO2003054166-A2.
XX
PD 03-JUL-2003.

PF	19-DEC-2002; 2002MO-US041225.
XX	
PR	20-DEC-2001; 2001US-0342603P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Jones KA, Schafer A;
DR	WPI, 2003-559141/52.
XX	
PT	Determining susceptibility of an individual to joint space narrowing,
PT	osteohipyte development and/or joint pain comprises identifying whether
PT	the individual has at least one polymorphism in a polynucleotide encoding
PT	a protein.
XX	
PS	Disclosure; SEQ ID NO 494; 297pp; English.
XX	
CC	The invention relates to a method of determining susceptibility of an
CC	individual to joint space narrowing and/or osteohipyte development and/or
CC	joint pain, comprising identifying whether the individual has at least one
CC	polymorphism in a polynucleotide encoding at least one of the protein
CC	listed in the specification. The methods, composition and agent are
CC	useful for modulating the susceptibility of an individual to joint space
CC	narrowing and/or osteohipyte development and/or joint pain that is
CC	associated with a disease, preferably osteoarthritis. The cell line and
CC	the non-human animal are useful for screening for an agent for diagnosing
CC	an individual having susceptibility to joint space narrowing and/or
CC	osteohipyte development and/or joint pain. This sequence corresponds to
CC	the polynucleotide encoding a protein listed in the specification. (Note:
CC	The sequence data for this patent did not form part of the printed
CC	specification but was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences).
XX	
SQ	Sequence 160198 BP; 48087 A; 32398 C; 32827 G; 46852 T; 0 U; 34 Other;
QY	Query Match 79.2%; Score 19.8; DB 10; Length 160198;
	Best Local Similarity 91.3%; Pred. No. 6.5e+02;
	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db	2 AAAAAAAAAACCTTCATCATTTTAA 24
	89087 AAAAAACAACCTTCATTTTAA 89065
RESULT 35	
ID	ABZ17257 standard; DNA; 2000 BP.
XX	
AC	ABZ17257;
XX	
DT	21-JAN-2003 (first entry)
XX	
DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 5062.
XX	
KW	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200216655-A2.
XX	
PD	28-FEB-2002.
XX	
PE	24-AUG-2001; 2001WO-US026685.
XX	
PR	24-AUG-2000; 2000US-0227866P.
XX	
PR	26-JAN-2001; 2001US-0264647P.
XX	
PR	22-JUN-2001; 2001US-0300111P.
XX	
PA	(SCRI) SCRIPS RES INST.
XX	
PA	(SYGN) SYNGENTA PARTICIPATIONS AG.
XX	
PI	Harper JF, Kreps J, Wang X, Zhu T;
XX	

DR	WPI, 2002-304127/34.
PT	Identifying a stress condition to which a plant cell has been exposed and
XX	producing plants with increased tolerance to these abiotic stresses.
PS	Claim 144; SEQ ID NO 5062; 577bp + Sequence Listing; English.
CC	The invention relates to identifying a stress condition to which a plant
CC	cell has been exposed, comprising: (a) contacting nucleic acid
CC	representative of expressed polynucleotides in the plant cell with an
CC	array or probes representative of the plant cell genome; and (b)
CC	detecting a profile of expressed polynucleotides in the plant cell
CC	characteristic of a stress response. The method is useful in the
CC	production of transgenic plants, cells and seeds and in producing plants
CC	with increased tolerance to abiotic stress. The present sequence is that
CC	of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
CC	in methods of the invention. Note: the sequence data for this patent is
CC	not represented in the printed specification but is based on sequence
CC	information supplied to Derwent by the European Patent Office
SQ	Sequence 2000 BP; 699 A; 319 C; 314 G; 668 T; 0 U; 0 Other;
Query Match	77.6%; Score 19.4; DB 6; Length 2000;
Beet Local Similarity	95.2%; Pred. No. 8.4e+02;
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY	1 AAAAAAAAACTCATCATT 21
Db	1051 AAAAAAAAACTCATCATT 1071
RESULT 36	
AEB76936	
ID	AEB76936 standard; DNA; 68295 BP.
AC	AEB76936;
XX	
DT	20-OCT-2005 (first entry)
XX	
DE	Chicken ovomucoid gene controlling region, OM24 SEQ ID NO: 36 #2.
XX	
KM	Gene expression; pharmaceutical; transgenic animal; diagnostic;
KW	feed-additive; ovomucoid; gene; ds.
XX	
OS	Gallus gallus.
XX	
FH	Key
FT	misc_feature
FT	Location/Qualifiers
FT	1..68295
FT	/tag= a
FT	/note= "Ovomucoid region"
FT	26416..36390
FT	/tag= b
FT	/note= "Ovomucoid gene expression controlling region,
FT	fragment A"
FT	32364..36299
FT	/tag= c
FT	/note= "This region is specifically claimed in claim 12"
FT	32365..36299
FT	/tag= d
FT	/note= "Ovomucoid gene expression controlling region,
FT	fragment B"
FT	34473..36248
FT	/tag= e
FT	/note= "This region is specifically claimed in claim 12"
FT	34474..36248
FT	/tag= f
FT	/note= "Ovomucoid gene expression controlling region,
FT	fragment C"
XX	
PX	US2005176047-A1.
PD	11-AUG-2005.

XX 31-JAN-2005; 2005US-00047184.
 PF 30-NOV-2001; 2001US-0098716.
 XX 02-DEC-2002; 2002MO-US038413.
 PR 06-JUN-2003; 2003US-0476596P.
 PR 24-SEP-2003; 2003US-0505562P.
 PR 06-OCT-2003; 2003US-0509122P.
 PR 01-MAR-2004; 2004US-00790455.
 PR 21-MAY-2004; 2004US-00496731.
 PR 28-MAY-2004; 2004US-00856218.
 XX (HARV/) HARVEY A J.
 PA (LEAV/) LEAVITT M C.
 PA (WANG/) WANG Y.
 XX Harvey AJ, Leavitt MC, Wang Y;
 PI WPI; 2005-541761/55.
 DR WPI; 2005-541761/55.
 XX New avian ovomucoid gene expression controlling region, useful for
 PT expressing a wide range of desired proteins, e.g. proteins used as human
 PT and animal pharmaceuticals, diagnostics, and livestock feed additives.
 PS Claim 9; SEQ ID NO 36, 158bp; English.
 XX The present invention relates to a nucleic acid molecule comprising an
 CC ovomucoid (OM) gene expression controlling region isolated from a chicken
 CC and/or operably linked to a coding sequence encoding an amino acid
 CC expressing other than ovomucoid. The nucleic acid molecule is useful for
 CC expressing, in large and at low cost, a wide range of desired proteins and
 CC including those used as human and animal pharmaceuticals, diagnostics and
 CC livestock feed additives or other proteins, e.g. growth hormones,
 CC cytokines, structural proteins, or enzymes. The invention is also useful
 CC for producing transgenic animal. The present sequence is the chicken
 CC ovomucoid gene controlling region comprising ovinhibitor and ovomucoid
 CC gene. Note: The present sequence is the SEQ ID NO: 36 which is shown on
 CC Figure 14 of the disclosure. This sequence differs from the SEQ ID NO: 36
 CC given in the sequence listing (see AEB76922).
 XX Sequence 68295 BP; 16921 A; 15748 C; 17034 G; 18592 T; 0 U; 0 Other;
 SQ
 Query Match 77.6%; Score 19.4; DB 14; Length 68295;
 Best Local Similarity 95.2%; Pred. No. 9.1e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAACCTTCATCATTT 21
 Db 45769 AAAAAAAAAACCTTCATCATTT 45789
 RESULT 37
 ADM44335
 ID ADM44335 standard; DNA; 75815 BP.
 XX ADM44335;
 AC
 XX 24-MAR-2005 (first entry)
 DT
 XX
 DE Chicken OMC24 clone DNA.
 XX Artificial chromosome; promoter; gene expression; transgenic animal;
 KM antibody production; cytokine; pharmaceutical; ds; ovomucoid; OM.
 XX Gallus gallus.
 OS
 XX US2005003414-A1.
 PN
 XX 06-JAN-2005.
 PD
 XX 28-MAY-2004; 2004US-00856218.
 PF
 XX 30-NOV-2001; 2001US-0098716.

PR 02-DEC-2002; 2002MO-US038413.
 PR 06-JUN-2003; 2003US-0476596P.
 PR 24-SEP-2003; 2003US-0505562P.
 PR 06-OCT-2003; 2003US-0509122P.
 PR 01-MAR-2004; 2004US-00790455.
 PR 21-MAY-2004; 2004US-00496731.
 XX (HARV/) HARVEY A J.
 PA (LEAV/) LEAVITT M C.
 PA (WANG/) WANG Y.
 XX Harvey AJ, Leavitt MC, Wang Y;
 PI WPI; 2005-065235/07.
 DR WPI; 2005-065235/07.
 XX New ovomucoid promoter, useful for producing transgenic avian,
 PT pharmaceutical compositions, cytokines, and antibodies.
 PS Claim 8; SEQ ID NO 36; 132bp; English.
 XX The present invention relates to the nucleic acid molecule comprising an
 CC ovomucoid gene expression controlling region isolated from a chicken and
 CC a coding polypeptide other than ovomucoid. The invention is useful for
 CC producing transgenic avian and for producing biological products such as
 CC pharmaceutical compositions, biological molecules including cytokines,
 CC antibodies. The present sequence is the chicken OMC24 clone DNA. This
 CC sequence is used in the construction of ovomucoid promoter-bacterial
 CC artificial chromosome expression vector with an antibody heavy chain or
 CC antibody light chain coding sequence.
 XX Sequence 75815 BP; 19047 A; 17581 C; 18725 G; 20462 T; 0 U; 0 Other;
 SQ
 Query Match 77.6%; Score 19.4; DB 14; Length 75815;
 Best Local Similarity 95.2%; Pred. No. 9.1e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAACCTTCATCATTT 21
 Db 45769 AAAAAAAAAACCTTCATCATTT 45789
 RESULT 38
 AEA00969
 ID AEA00969 standard; DNA; 75815 BP.
 XX AEA00969;
 AC
 XX 14-JUL-2005 (first entry)
 DT
 XX DNA insert OMC24 with ovinhibitor and ovomucoid sequences.
 DE
 XX Promoter; gene expression; gene regulation; protein production;
 KM regulatory region; gene activation; animal disease model; ds; ovomucoid;
 KM ovinhibitor.
 XX Gallus gallus.
 OS
 XX WO2005040215-A2.
 PN
 XX 06-MAY-2005.
 PD
 XX 28-MAY-2004; 2004MO-US016827.
 PF
 XX 06-JUN-2003; 2003US-0476596P.
 PR 28-JUL-2003; 2003US-0490452P.
 PR 24-SEP-2003; 2003US-0505562P.
 PR 06-OCT-2003; 2003US-0509122P.
 PR 15-JAN-2004; 2004US-0536677P.
 PR 01-MAR-2004; 2004US-00790455.
 PR 21-MAY-2004; 2004US-00496731.
 XX (AVIG-) AVIGENICS INC.

PI Harvey AJ, Leavitt MC, Wang Y;
XX WPI; 2005-333493/34.
XX
PT New nucleic acid molecule having an ovomucoid gene expression controlling
PT region isolated from a chicken and/or a coding sequence encoding a
PT polypeptide other than ovomucoid, useful for producing models for various
PT diseases in humans.
XX
XX Example 6; SEQ ID NO 36; 212pp; English.
PS
XX The present invention relates to an avian, preferably chicken, ovomucoid
CC gene expression control region, which is useful for the expression of
CC nucleotide sequences encoding a protein of interest in a transfected
CC avian cell, such as an oviduct cell. The chicken ovomucoid gene is highly
CC expressed in the tubular glands of the mature hen oviduct and represents
CC a suitable candidate for an efficient promoter for heterologous protein
CC production in transgenic animals. The regulatory region of the ovomucoid
CC locus may extend over a nucleic acid region of about 10kb of DNA 5'
CC upstream of the transcription start site and includes a CRI element.
CC Provided is a nucleic acid molecule comprising an ovomucoid gene
CC expression controlling region isolated from a chicken and/or a coding
CC sequence encoding a polypeptide other than ovomucoid. Specifically, the
CC ovomucoid gene expression control region has a sequence given in SEQ ID
CC NO:26 (see AEA00959), or is the avian nucleic acid contained in SEQ ID
CC NO:36, (see AEA00959). The polypeptide is present in egg white produced
CC by a transgenic avian. The polypeptide is a fusion protein or a CRI4-Fc
CC fusion protein. The methods and compositions of the present invention are
CC useful in understanding the action of a single gene in the context of the
CC whole animal and the phenomena of gene activation, expression and
CC interaction, and for producing models for various diseases in humans and
CC other animals, and in the study of genetics, including genetic mechanisms
CC and functions. The present sequence is a DNA insert from a BAC clone
CC encompassing the ovoinhibitor and ovomucoid genes, designated OMC24.
XX
SQ Sequence 75815 BP; 19047 A; 17581 C; 18725 G; 20462 T; 0 U; 0 Other;
XX
Query Match 77.6%; Score 19.4; DB 14; Length 75815;
Best Local Similarity 95.2%; Pred. No. 9.1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCTTCATCTT 21
Db 45769 AAAAAAAAAACCTTCAGCATTT 45789
XX
RESULT 39
AEB76922
ID AEB76922 standard; DNA; 75815 BP.
XX
AC AEB76922;
XX
DT 20-OCT-2005 (first entry)
XX
DE Chicken ovomucoid gene controlling region, OMC24 SEQ ID NO: 36 #1.
XX
KM Gene expression; pharmaceutical; transgenic animal; diagnostic;
KM feed-additive; ovomucoid; gene; ds.
XX
OS Gallus gallus.
XX
XX Key Location/Qualifiers
FH 1. .68295
FT misc_feature
FT /cag= a
FT /note= "Ovomucoid region"
FT 26416. .36390
FT misc_feature
FT /cag= b
FT /note= "Ovomucoid gene expression controlling region,
FT fragment A"
FT /note= "This region is specifically claimed in claim 12"
FT 32364. .36299
FT misc_feature
FT /cag= c
FT /note= "This region is specifically claimed in claim 12"

FT misc_feature 32365. .36299
FT /cag= d
FT /note= "Ovomucoid gene expression controlling region,
FT fragment B"
FT 34473. .36248
FT misc_feature
FT /cag= e
FT /note= "This region is specifically claimed in claim 12"
FT 34474. .36248
FT misc_feature
FT /cag= f
FT /note= "Ovomucoid gene expression controlling region,
FT fragment C"
FT 68296. .75815
FT misc_feature
FT /cag= g
FT /note= "BAC vector"
XX
XX US2005176047-A1.
XX
XX 11-AUG-2005.
XX
XX 31-JAN-2005; 2005US-00047184.
XX
XX 30-NOV-2001; 2001US-0098716.
XX 02-DEC-2002; 2002MO-US038413.
XX 06-JUN-2003; 2003US-0476596P.
XX 24-SEP-2003; 2003US-0505562P.
XX 06-OCT-2003; 2003US-0509122P.
XX 01-MAR-2004; 2004US-00790455.
XX 21-MAY-2004; 2004US-00496731.
XX 28-MAY-2004; 2004US-00856218.
XX
XX (HARV/) HARVEY A J.
XX (LEAV/) LEAVITT M C.
XX (WANG/) WANG Y.
XX
PI Harvey AJ, Leavitt MC, Wang Y;
XX
XX WPI; 2005-541761/55.
XX
XX New avian ovomucoid gene expression controlling region, useful for
PT expressing a wide range of desired proteins, e.g. proteins used as human
PT and animal pharmaceuticals, diagnostics, and livestock feed additives.
XX
PS Claim 9; SEQ ID NO 36; 158pp; English.
XX
XX The present invention relates to a nucleic acid molecule comprising an
CC ovomucoid (OM) gene expression controlling region isolated from a chicken
CC and/or operably linked to a coding sequence encoding an amino acid
CC sequence other than ovomucoid. The nucleic acid molecule is useful for
CC expressing, in large and at low cost, a wide range of desired proteins
CC including those used as human and animal pharmaceuticals, diagnostics and
CC livestock feed additives or other proteins, e.g. growth hormones,
CC cytokines, structural proteins, or enzymes. The invention is also useful
CC for producing transgenic animal. The present sequence is the chicken
CC ovomucoid gene controlling region comprising ovoinhibitor and ovomucoid
CC gene. Note: The present sequence is the SEQ ID NO: 36 which is given in
CC the sequence listing. This sequence differs from the SEQ ID NO: 36 shown
CC on Figure 14 of the disclosure (see AEB76936).
XX
SQ Sequence 75815 BP; 19047 A; 17581 C; 18725 G; 20462 T; 0 U; 0 Other;
XX
Query Match 77.6%; Score 19.4; DB 14; Length 75815;
Best Local Similarity 95.2%; Pred. No. 9.1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCTTCATCTT 21
Db 45769 AAAAAAAAAACCTTCAGCATTT 45789
XX
RESULT 40
ADM44343
ID ADM44343 standard; DNA; 77872 BP.
XX

AC ADW44343;
XX
XX 24-MAR-2005 (first entry)
XX
XX Ovomucoid promoter-bacterial artificial chromosome expression vector.
DE
XX Artificial chromosome; promoter; gene expression; transgenic animal;
XX antibody production; cytokine; pharmaceutical; ds; ovomucoid; OM; CTLA4;
KM cytotoxic T lymphocyte antigen 4.
XX
XX Gallus gallus.
OS Homo sapiens.
OS Chimeric.
OS Unidentified.
XX US2005003414-A1.
PN
XX 06-JAN-2005.
PD
XX 28-MAY-2004; 2004US-00856218.
PF
XX 30-NOV-2001; 2001US-00998716.
PR 02-DEC-2002; 2002WO-05038413.
PR 06-JUN-2003; 2003US-0476596P.
PR 24-SEP-2003; 2003US-0505562P.
PR 06-OCT-2003; 2003US-0509122P.
PR 01-MAR-2004; 2004US-00790455.
PR 21-MAY-2004; 2004US-00496731.
XX
XX (HARV/) HARVEY A J.
PA (LEAV/) LEAVITT M C.
PA (WANG/) WANG Y.
XX
XX Harvey AJ, Leavitt MC, Wang Y;
PI
XX WPI; 2005-065235/07.
DR
XX
XX New ovomucoid promoter, useful for producing transgenic avian,
PT pharmaceutical compositions, cytokines, and antibodies.
XX
XX Example 12; SEQ ID NO 44; 132pp; English.
PS
XX The present invention relates to the nucleic acid molecule comprising an
CC ovomucoid gene expression controlling region isolated from a chicken and
CC a coding polypeptide other than ovomucoid. The invention is useful for
CC producing transgenic avian and for producing biological products such as
CC pharmaceutical compositions, biological molecules including cytokines,
CC antibodies. The present sequence is the ovomucoid promoter-bacterial
CC artificial chromosome expression vector with a CTLA-Fc fusion coding
CC sequence and an atb site.
XX
SQ Sequence 77872 BP; 19512 A; 18197 C; 19302 G; 20861 T; 0 U; 0 Other;
SQ
Query Match 77.6%; Score 19.4; DB 14; Length 77872;
Best Local Similarity 95.2%; Pred. No. 9.1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCTTCATCATTT 21
DB 4142 AAAAAAAAAACCTTCAGCATTT 4162
RESULT 41
AEA00977
ID AEA00977 standard; DNA; 77872 BP.
XX
XX AEA00977;
AC
XX 14-JUL-2005 (first entry)
DT
XX
XX Ovomucoid promoter-BAC vector with CTLA4-Fc fusion sequence.
DE
XX Promoter; gene expression; gene regulation; protein production;
KM

KM regulatory region; gene activation; animal disease model; ds;
KM ovomucoid promoter; vector.
XX
XX
XX Gallus gallus.
OS Synthetic.
OS WO2005040215-A2.
PN
XX 06-MAY-2005.
PD
XX 28-MAY-2004; 2004WO-US016827.
PF
XX 06-JUN-2003; 2003US-0476596P.
PR 28-JUL-2003; 2003US-0490452P.
PR 24-SEP-2003; 2003US-0505562P.
PR 06-OCT-2003; 2003US-0509122P.
PR 15-JAN-2004; 2004US-0536677P.
PR 01-MAR-2004; 2004US-00790455.
PR 21-MAY-2004; 2004US-00496731.
XX
XX (AVIG-) AVIGENICS INC.
PA
XX Harvey AJ, Leavitt MC, Wang Y;
PI
XX WPI; 2005-333493/34.
DR
XX
XX New nucleic acid molecule having an ovomucoid gene expression controlling
PT region isolated from a chicken and/or a coding sequence encoding a
PT polypeptide other than ovomucoid, useful for producing models for various
PT diseases in humans.
XX
XX Example 12; SEQ ID NO 44; 212pp; English.
PS
XX
XX The present invention relates to an avian, preferably chicken, ovomucoid
CC gene expression control region, which is useful for the expression of
CC nucleotide sequences encoding a protein of interest in a transfected
CC avian cell, such as an oviduct cell. The chicken ovomucoid gene is highly
CC expressed in the tubular glands of the mature hen oviduct and represents
CC a suitable candidate for an efficient promoter for heterologous protein
CC production in transgenic animals. The regulatory region of the ovomucoid
CC locus may extend over a nucleic acid region of about 10kb of DNA 5'
CC upstream of the transcription start site and includes a CRI element.
CC Provided is a nucleic acid molecule comprising an ovomucoid gene
CC expression controlling region isolated from a chicken and/or a coding
CC sequence encoding a polypeptide other than ovomucoid. Specifically, the
CC ovomucoid gene expression control region has a sequence given in SEQ ID
CC NO:26 (see AEA00969) or is the avian nucleic acid contained in SEQ ID
CC NO:36, (see AEA00969). The polypeptide is present in egg white produced
CC by a transgenic avian. The polypeptide is a fusion protein or a CTLA4-Fc
CC fusion protein. The methods and compositions of the present invention are
CC useful in understanding the action of a single gene in the context of the
CC whole animal and the phenomena of gene activation, expression and
CC interaction, and for producing models for various diseases in humans and
CC other animals, and in the study of genetics, including genetic mechanisms
CC and functions. The present sequence is DNA of an ovomucoid gene
CC expression controlling region-BAC expression vector with a CTLA4-Fc
CC fusion coding sequence and an atb site. This was constructed using
CC nucleotide sequences for an immunoglobulin constant region (IgG1Fc).
XX
SQ Sequence 77872 BP; 19512 A; 18197 C; 19302 G; 20861 T; 0 U; 0 Other;
SQ
Query Match 77.6%; Score 19.4; DB 14; Length 77872;
Best Local Similarity 95.2%; Pred. No. 9.1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAACCTTCATCATTT 21
DB 4142 AAAAAAAAAACCTTCAGCATTT 4162
RESULT 42
AEB76930
ID AEB76930 standard; DNA; 77872 BP.


```
XX ABB76930;
AC
XX
XX 20-OCT-2005 (first entry)
DT
XX
XX
DE Chicken OMC24-actB-IRES-CTLA4-IgG1 Fc fusion DNA, SEQ ID NO: 44.
XX
XX Gene expression; pharmaceutical; transgenic animal; diagnostic;
KM feed-additive; ovomucoid; gene fusion; internal ribosome entry site; IgG;
XX immunoglobulin; cytotoxic T lymphocyte antigen 4 receptor; ds.
XX
OS Gallus gallus.
OS Homo sapiens.
OS Chimeric.
OS Unidentified.
XX
XX US2005176047-A1.
PN
XX
XX 11-AUG-2005.
PD
XX
XX 31-JAN-2005; 2005US-00047184.
PF
XX
XX 30-NOV-2001; 2001US-00998716.
PR 02-DEC-2002; 2002MO-US038413.
PR 06-JUN-2003; 2003US-0476596P.
PR 24-SEP-2003; 2003US-0505562P.
PR 06-OCT-2003; 2003US-0509122P.
PR 01-MAR-2004; 2004US-00790455.
PR 21-MAY-2004; 2004US-00496731.
PR 28-MAY-2004; 2004US-00856218.
XX
XX (HARV/) HARVEY A J.
PA (LEAV/) LEAVITT M C.
PA (WANG/) WANG Y.
XX
XX Harvey AJ, Leavitt MC, Wang Y;
PI
XX
XX WPI; 2005-541761/55.
DR
XX
XX New avian ovomucoid gene expression controlling region, useful for
PT expressing a wide range of desired proteins, e.g. proteins used as human
PT and animal pharmaceuticals, diagnostics, and livestock feed additives.
XX
XX Example 12; SEQ ID NO 44; 158bp; English.
PS
XX
XX The present invention relates to a nucleic acid molecule comprising an
CC ovomucoid (OM) gene expression controlling region isolated from a chicken
CC and/or operably linked to a coding sequence encoding an amino acid
CC sequence other than ovomucoid. The nucleic acid molecule is useful for
CC expressing, in large and at low cost, a wide range of desired proteins
CC including those used as human and animal pharmaceuticals, diagnostics and
CC livestock feed additives or other proteins, e.g. growth hormones,
CC cytokines, structural proteins, or enzymes. The invention is also useful
CC for producing transgenic animal. The present sequence is a fusion DNA
CC comprising chicken ovomucoid gene controlling region, OMC24, internal
CC ribosome entry site (IRES) DNA, actB site, extra cellular domain of
CC cytotoxic T lymphocyte antigen 4 (CTLA4) receptor protein DNA and human
CC IgG1 constant region (Fc).
XX
XX
SQ Sequence 77872 BP; 19512 A; 18197 C; 19302 G; 20861 T; 0 U; 0 Other;
XX
XX
XX Query Match 77.6%; Score 19.4; DB 14; Length 77872;
XX Best Local Similarity 95.2%; Pred. No. 9.1e+02;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAACCTTCATCATTT 21
QY |||||||||||||||||||
DB 4142 AAAAAAAAAACCTTCAGCATTT 4162
XX |||||||||||||||||||
XX
XX
XX RESULT 43
XX AAH37044/c
XX AAH37044 standard; cDNA; 151 BP.
```

```
XX
XX AAH37044;
AC
XX
XX 03-SEP-2001 (first entry)
DT
XX
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:4126.
XX
XX Human colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200122920-A2.
PN
XX
XX 05-APR-2001.
PD
XX
XX 28-SEP-2000; 2000MO-US026524.
PF
XX
XX 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX
XX WPI; 2001-235357/24.
DR
XX
XX P-PSDB; AAG77637.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX Claim 1; Page 5980-5981; 9803pp; English.
PS
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC to 1052, 7921 and 7922
XX
XX
SQ Sequence 151 BP; 40 A; 33 C; 29 G; 42 T; 0 U; 7 Other;
XX
XX
XX Query Match 76.8%; Score 19.2; DB 4; Length 151;
XX Best Local Similarity 87.5%; Pred. No. 9.4e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 AAAAAAAAAACCTTCATCATTTAA 24
QY |||||||||||||||||||
DB 103 AAAAAAAAAACCTTTTAATTAA 80
XX |||||||||||||||||||
XX
XX
XX RESULT 44
XX ABX54379
XX ID ABX54379 standard; cDNA; 247 BP.
XX
XX ABX54379;
AC
XX
XX 25-FEB-2003 (first entry)
DT
XX
XX Bovine EST associated with lactation/muscle/fat deposition #4308.
DE
XX
```

KM Bovine; ss; EST; expressed sequence tag; laccation; LMPD;
 KM muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 PN US2002137160-A1.
 PD
 PD 26-SEP-2002.
 XX
 PF 26-OCT-2001; 2001US-00983965.
 XX
 PR 17-DEC-1998; 98US-0113678P.
 PR 15-DEC-1999; 99US-00465231.
 XX
 PA (BYATT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;
 DR WPI; 2003-102386/09.
 XX
 PT Purified nucleic acid molecules, useful for genome mapping, gene
 PT identification and analysis, cattle breeding or preparation of constructs
 PT for cattle gene expression and genetically improved cattle.
 XX
 PS Claim 2; SEQ ID NO 4308; 38pp; English.
 XX
 CC The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMPD), derived from
 CC cattle, and the LMPD nucleic acid can specifically hybridize to a second
 CC nucleic acid molecule comprising any of 5912 nucleotide sequences,
 CC appearing as ABX50072-ABX55983, or complements of them. Also included are
 CC (1) a transformed cell having a nucleic acid comprising an LMPD nucleic
 CC acid linked to a promoter and a 3' non translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 5912 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridization between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMPD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the 5912
 CC bovine LMPD EST (expressed sequence tag) nucleic acids. Note: The present
 CC sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?docID=20020137160
 XX
 SQ Sequence 247 BP; 99 A; 32 C; 40 G; 76 T; 0 U; 0 Other;
 Query Match 76.8%; Score 19.2; DB 8; Length 247;
 Best Local Similarity 87.5%; Pred. No. 9.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
 Db 115 AAAAAAAAAACGACATCATTTATA 138

RESULT 45
 ACF85962
 ID ACF85962 standard; DNA; 500 BP.
 XX
 AC ACF85962;

XX
 DT 02-JUN-2005 (first entry)
 XX
 DE Human SIRS/sepsis diagnostic marker DNA fragment 4822.
 XX
 KW Systemic inflammatory response syndrome; SIRS; antibacterial;
 KW immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.
 OS Homo sapiens.
 XX
 PN WO2004087949-A2.
 PD
 PD 14-OCT-2004.
 XX
 PF 31-MAR-2004; 2004WO-EP003419.
 XX
 PR 02-APR-2003; 2003DE-01015031.
 PR 08-AUG-2003; 2003DE-01036511.
 PR 02-SEP-2003; 2003DE-01040395.
 XX
 PA (SIRS-) SIRS LAB GMBH.
 PI
 PI Russwurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Delinger H;
 DR WPI; 2004-748070/73.
 XX
 PT In vitro detection of systemic inflammatory response syndrome and related
 PT conditions, for e.g. monitoring progression, comprises detecting abnormal
 PT expression of disease-related genes.
 XX
 PS Disclosure, Page; 75pp; German.
 XX
 CC The invention relates to a novel method for in vitro detection of
 CC systemic inflammatory response syndrome (SIRS). The method comprises
 CC detecting abnormal expression of disease-related genes, or their
 CC associated peptides. The method of the invention demonstrates
 CC antibacterial, immunosuppressive and antiinflammatory applications and
 CC may be used for early differential diagnosis, monitoring progression,
 CC assessing risk, assessing the likely response to treatment and for post
 CC mortem diagnosis of systemic inflammatory response syndrome, sepsis and
 CC sepsis-like conditions. The recombinant or synthetic nucleic acid
 CC sequences of the invention, or derived proteins or peptides, may be
 CC useful as calibrants in assays for the specified diseases, for evaluating
 CC activity or toxicity in screening for active agents and/or for
 CC preparation of agents for treatment or prevention of the specified
 CC diseases. The current sequence is that of a human SIRS/sepsis diagnostic
 CC marker DNA fragment of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at fcp.wipo.int/pub/published
 CC pct sequences. Furthermore, a number of arbitrary SEQ ID NO.s are
 CC disclosed within the specification, however, these have not been taken
 CC into account during indexing due to inconsistencies in application and
 CC format
 XX
 SQ Sequence 500 BP; 178 A; 104 C; 74 G; 141 T; 0 U; 3 Other;
 Query Match 76.8%; Score 19.2; DB 13; Length 500;
 Best Local Similarity 87.5%; Pred. No. 9.6e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
 Db 322 AAAAAAAAACTTTAATCATTTAA 345

RESULT 46
 ABA62042/C
 ID ABA62042 standard; DNA; 541 BP.
 XX
 AC ABA62042;
 XX
 DT 01-FEB-2002 (first entry)
 XX

DE Human foetal liver single exon nucleic acid probe #10347.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000669.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human fetal liver.
 XX
 PS Claim 1; SEQ ID NO 10347; 639bp + Sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human fetal liver. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 541 BP; 178 A; 93 C; 80 G; 190 T; 0 U; 0 Other;
 XX
 Query Match 76.8%; Score 19.2; DB 4; Length 541;
 Best Local Similarity 87.5%; Pred. No. 9.7e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAACCTTCATCATTTAA 24
 Db 425 AAAAAAAAAACCTTCATCATTTAA 402
 XX
 RESULT 47
 AA11981/c
 ID AA11981 standard; DNA; 541 BP.
 XX
 AC AA11981;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #10667 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human placenta.
 XX
 PS Claim 25; SEQ ID NO 10667; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders
 CC
 SQ Sequence 541 BP; 178 A; 93 C; 80 G; 190 T; 0 U; 0 Other;
 XX
 Query Match 76.8%; Score 19.2; DB 4; Length 541;
 Best Local Similarity 87.5%; Pred. No. 9.7e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAACCTTCATCATTTAA 24
 Db 425 AAAAAAAAAACCTTCATCATTTAA 402
 XX
 RESULT 48
 ABQ39865
 ID ABQ39865 standard; DNA; 584 BP.
 XX
 AC ABQ39865;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26456.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 DE 01-SEP-2001; 2001WO-EP010074.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPig-) EPiGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 DR WPI; 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 PS Claim 12; 56bp + Sequence listing; 56bp; German.

```

CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ3410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX
S0 Sequence 584 BP; 102 A; 63 C; 169 G; 230 T; 0 U; 0 Other;

Query Match          76.8%; Score 19.2; DB 6; Length 584;
Beet Local Similarity 87.5%; Pred. No. 9.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy      2  AAAAAAAAACTGTCATCTTAAAC 25
Db      225 AAAAAAAAACTGTCATCTTAAAC 202

RESULT 50
AC AAA98852
AC AAA98852;
XX
XX ID AAA98852 standard; cDNA; 1620 BP.
XX
DE Human differentiation-associated endothelial EST cDNA SEQ ID NO 28.
XX
XX EST: expressed sequence tag; human; cell differentiation; antidiabetic;
XX cell proliferation; endothelial tissue; angiogenic; antiproliferatic;
XX optineurological; nephrotropic; cystostatic; hepatotropic; antiallergic;
XX antiarteriosclerotic; antirheumatic; gene therapy; angiogenic disease;
XX treatment; rheumatoid arthritis; haemangioma; angiofibroma; eye disease;
XX diabetic retinopathy; glaucoma; kidney disease; glomerulonephritis;
XX thrombotic microangiopathic syndrome; glomerulopathy; fibrotic disease;
XX mesangial cell proliferative disease; ss.
XX
XX Homo sapiens.
XX
XX DE19911684-A1.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-1999; 99DE-01011684.
XX 09-MAR-1999; 99DE-01011684.
XX 01-OCT-1999; 99DE-01048679.
XX
XX (SCHD ) SCHERING AG.
XX
XX Glienke J, Thierauch K, Hinzmann B, Pilarsky C;
XX WPI; 2000-572267/53.
XX
XX Nucleic acid sequences from human endothelial cells, useful for gene
XX therapy of angiogenesis and for identifying antiangiogenic agents.
XX
XX Claim 1a; Page 51; 84pp; German.
XX
XX This invention describes novel human nucleic acid sequences (i) from
XX endothelial cells which have angiogenic, antiproliferatic, antidiabetic.
XX

```

CC ophthalmological, nephrotropic, cytostatic, hepatotropic,
 CC antiarteriosclerotic, antiarthritic, antirheumatic activity and can be
 CC used for gene therapy. The nucleic acid sequences may find use as tools
 CC to identify agents against angiogenic diseases. The polypeptides may be
 CC used in medicaments for gene therapy to treat angiogenic diseases. (1) in
 CC sense or antisense form may be used. The genomic genes may also be used
 CC with suitable promoters and/or enhancers. The sequences (nucleic acid and
 CC proteins) may be used to treat psoriasis, arthritis, especially
 CC rheumatoid arthritis, haemangioma, angiodroma, eye diseases, especially
 CC diabetic retinopathy, neovascular glaucoma, kidney disease, such as
 CC glomerulonephritis, diabetic nephropathy, malign nephrosclerosis,
 CC thrombotic microangiopathic syndrome, transplantation rejection and
 CC glomerulopathy, fibrotic disease, such as liver cirrhosis, mesangial cell
 CC proliferative disease, arteriosclerosis and injury to nerve tissue
 XX

Sequence 1620 BP; 504 A; 312 C; 266 G; 538 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 3; Length 1620;
 Best Local Similarity 87.5%; Pred. No. 9.9e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
 |||||
 Db 360 AAAAAAAAAACCTTTTACATTTAA 383

Search completed: December 14, 2005, 02:43:15
 Job time : 208.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:34:03 ; Search time 1752.1 Seconds

(without alignments)
667.586 Million cell updates/sec

Title: US-10-681-773-9

Perfect score: 25

Sequence: 1 aaaaaaaaaacattcattcaaac 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339354128 residues 82156650

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :
EST:
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gest1.*
10: gb_gest2.*
11: gb_gest3.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	22.4	89.6	382 5	BM193564 BM193564
2	22.4	89.6	461 7	CV014642 CV014642
3	22.4	89.6	484 10	CL378878 CL378878
4	21.8	87.2	368 5	BM066980 BM066980
5	21.8	87.2	566 10	BX158115 BX158115
6	21.8	87.2	976 10	CZ935905 CZ935905
7	21.4	85.6	537 7	CV014040 CV014040
8	21.4	85.6	545 7	CV014702 CV014702
9	21	84.0	234 1	AV024200 AV024200
10	21	84.0	342 2	BG736368 BG736368
11	21	84.0	344 2	AQ096468 AQ096468
12	21	84.0	421 2	BR447256 BR447256
13	21	84.0	650 2	BG583394 BG583394
14	20.8	83.2	299 8	DN340148 DN340148
15	20.8	83.2	308 2	BG012600 BG012600
16	20.8	83.2	373 11	CR876225 CR876225
17	20.8	83.2	398 7	CO349733 CO349733
18	20.8	83.2	427 2	BF654733 BF654733
19	20.8	83.2	454 1	AI620206 AI620206
20	20.8	83.2	516 11	DR32977T DR32977T
21	20.8	83.2	547 1	AV765288 AV765288
22	20.8	83.2	582 7	CO911063 CO911063

C 23	20.8	83.2	586 10	AG242106 AG242106
C 24	20.8	83.2	587 3	B1961316 B1961316
C 25	20.8	83.2	607 7	CV014565 CV014565
C 26	20.8	83.2	643 9	BH765063 BH765063
C 27	20.8	83.2	661 1	BB088272 BB088272
C 28	20.8	83.2	676 10	AG314967 AG314967
C 29	20.8	83.2	687 3	B1451869 B1451869
C 30	20.8	83.2	696 10	CE502891 CE502891
C 31	20.8	83.2	739 9	BH567772 BH567772
C 32	20.8	83.2	740 7	CK362049 CK362049
C 33	20.8	83.2	761 7	CO253171 CO253171
C 34	20.8	83.2	764 8	DN527189 DN527189
C 35	20.8	83.2	773 3	CC573268 CC573268
C 36	20.8	83.2	794 10	AG316741 AG316741
C 37	20.8	83.2	837 10	DU088441 DU088441
C 38	20.8	83.2	839 8	DR882998 DR882998
C 39	20.8	83.2	864 8	DN524442 DN524442
C 40	20.8	83.2	881 10	CZ527449 CZ527449
C 41	20.8	83.2	891 9	AZ533375 AZ533375
C 42	20.8	83.2	894 6	CF579446 CF579446
C 43	20.8	83.2	938 10	CZ933563 CZ933563
C 44	20.8	83.2	981 5	BH167817 BH167817
C 45	20.8	83.2	1014 9	BH157390 BH157390
C 46	20.8	83.2	1101 10	CNS006X CNS006X
C 47	20.8	83.2	1161 9	CC257387 CC257387
C 48	20.8	83.2	1177 10	CL039316 CL039316
C 49	20.8	83.2	1457 9	CC230460 CC230460
C 50	20.8	83.2	1457 9	CC230460 CC230460
C 51	20.4	81.6	269 5	BH756057 BH756057
C 52	20.4	81.6	328 10	CB474507 CB474507
C 53	20.4	81.6	424 1	AA139976 AA139976
C 54	20.4	81.6	445 3	BJ087699 BJ087699
C 55	20.4	81.6	462 7	CV014251 CV014251
C 56	20.4	81.6	511 7	CV014714 CV014714
C 57	20.4	81.6	512 7	CV014758 CV014758
C 58	20.4	81.6	526 10	AJ587484 AJ587484
C 59	20.4	81.6	544 8	CX514125 CX514125
C 60	20.4	81.6	567 8	CX514126 CX514126
C 61	20.4	81.6	575 7	CV066964 CV066964
C 62	20.4	81.6	580 7	CV013575 CV013575
C 63	20.4	81.6	592 8	CX470846 CX470846
C 64	20.4	81.6	609 9	CV066934 CV066934
C 65	20.4	81.6	618 5	BX782883 BX782883
C 66	20.4	81.6	630 8	CX829846 CX829846
C 67	20.4	81.6	637 3	BJ076303 BJ076303
C 68	20.4	81.6	637 7	CR586309 CR586309
C 69	20.4	81.6	641 7	CV014037 CV014037
C 70	20.4	81.6	740 5	BX781325 BX781325
C 71	20.4	81.6	770 7	CK573701 CK573701
C 72	20.4	81.6	830 6	CF151384 CF151384
C 73	20.4	81.6	831 5	BX774586 BX774586
C 74	20.4	81.6	840 7	CR414352 CR414352
C 75	20.4	81.6	849 10	CZ538249 CZ538249
C 76	20.2	80.8	893 5	BX757310 BX757310
C 77	20.2	80.8	912 10	BX193862 BX193862
C 78	20.2	80.8	216 8	DN217252 DN217252
C 79	20.2	80.8	350 6	CD736498 CD736498
C 80	20.2	80.8	402 9	CE249520 CE249520
C 81	20.2	80.8	429 1	AI654110 AI654110
C 82	20.2	80.8	433 7	CO540602 CO540602
C 83	20.2	80.8	439 3	BH538028 BH538028
C 84	20.2	80.8	459 10	CE413112 CE413112
C 85	20.2	80.8	474 9	BZ244082 BZ244082
C 86	20.2	80.8	476 9	AZ642274 AZ642274
C 87	20.2	80.8	480 10	AG317997 AG317997
C 88	20.2	80.8	490 1	AW782608 AW782608
C 89	20.2	80.8	502 2	BG059619 BG059619
C 90	20.2	80.8	512 2	BF025003 BF025003
C 91	20.2	80.8	512 9	CC024699 CC024699
C 92	20.2	80.8	512 10	CG914999 CG914999
C 93	20.2	80.8	532 10	BX247445 BX247445
C 94	20.2	80.8	545 7	CV633511 CV633511
C 95	20.2	80.8	558 9	B2181756 B2181756

96	20.2	80.8	563	9	BZ359827	
97	20.2	80.8	570	9	AQ144303	
98	20.2	80.8	571	3	BM632771	
99	20.2	80.8	573	3	BM077279	
100	20.2	80.8	576	7	CV037778	
101	20.2	80.8	580	7	CV638491	
102	20.2	80.8	593	3	BM648126	
103	20.2	80.8	593	3	BM430900	
104	20.2	80.8	594	11	CR349321	
105	20.2	80.8	597	8	CV945972	
106	20.2	80.8	642	8	CV250218	
107	20.2	80.8	642	10	BM164568	
108	20.2	80.8	668	8	CV018640	
109	20.2	80.8	668	10	AG943725	
110	20.2	80.8	674	5	BU249049	
111	20.2	80.8	679	10	CE442857	
112	20.2	80.8	679	9	AZ600813	
113	20.2	80.8	684	9	BZ014286	
114	20.2	80.8	687	6	CD829442	
115	20.2	80.8	687	9	BZ024737	
116	20.2	80.8	695	9	BH970835	
117	20.2	80.8	712	10	CZ177359	
118	20.2	80.8	715	1	AJ737833	
119	20.2	80.8	719	6	CD821252	
120	20.2	80.8	730	6	CA365660	
121	20.2	80.8	732	7	CV644856	
122	20.2	80.8	753	7	CN728730	
123	20.2	80.8	767	8	DR424709	
124	20.2	80.8	771	10	CL149554	
125	20.2	80.8	776	11	CR051897	
126	20.2	80.8	794	10	CG150491	
127	20.2	80.8	822	9	BH699846	
128	20.2	80.8	821	9	BZ792359	
129	20.2	80.8	939	9	BZ792355	
130	20.2	80.8	974	11	CNS06KRFQ	
131	20.2	80.8	985	9	CC274837	
132	20.2	80.8	1004	10	CL092940	
133	20.2	80.8	1125	11	CNS06LC3	
134	20.2	80.8	1214	8	DN702516	
135	20.2	80.8	1303	10	AJ855979	
136	20	80.0	248	2	BG737091	
137	19.8	79.2	218	1	A1925340	
138	19.8	79.2	232	1	BB199498	
139	19.8	79.2	234	2	BB200831	
140	19.8	79.2	259	5	BU778259	
141	19.8	79.2	282	1	BB065228	
142	19.8	79.2	288	9	BZ437143	
143	19.8	79.2	342	6	CB343667	
144	19.8	79.2	413	7	CV012843	
145	19.8	79.2	427	2	BE448976	
146	19.8	79.2	428	10	CE477371	
147	19.8	79.2	430	5	BY484229	
148	19.8	79.2	438	7	CV750168	
149	19.8	79.2	444	7	CV750079	
150	19.8	79.2	445	7	CN772178	

ALIGNMENTS

RESULT 1	BM193564	382 bp	mRNA	linear	EST 01-JUN-2005
LOCUS	BM193564				
DEFINITION	BM193564 Nori Satoh unpublished cDNA library, cleaving embryo cDNA				
ACCESSION	BM193564				
VERSION	BM193564.1				
KEYWORDS	GI:24607966				
SOURCE	EST				
ORGANISM	Ciona intestinalis				
REFERENCE	Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Ciona.				
	1 (bases 1 to 382)				

AUTHORS Satou, Y., Shin-I, T., Kohara, Y. and Satoh, N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
 source
 1..382
 /organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="cic1101d07"
 /tissue_type="whole body"
 /dev_stage="cleaving embryo"
 /clone_lib="Nori Satoh unpublished cDNA library, cleaving embryo"

ORIGIN

Query Match 89.6%; Score 22.4; DB 5; Length 382;
 Best Local Similarity 92.0%; Pred. No. 6e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTCATCTTAAC 25
 Db 213 AAAAAAAAACTCATCTTAAC 237

RESULT 2 CV014642 461 bp mRNA linear EST 19-AUG-2004
LOCUS TL026A02 young leaf cDNA library of tea plant (Camellia sinensis)
DEFINITION Camellia sinensis cDNA clone TL026A02 5', mRNA sequence.

ACCESSION CV014642
VERSION CV014642.1 GI:51453856
KEYWORDS EST.
SOURCE Camellia sinensis (black tea)
ORGANISM Camellia sinensis

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; Ericales; Theaceae; Camellia.
 1 (bases 1 to 461)

AUTHORS Chen, L., Zhao, L.P. and Gao, Q.K.
TITLE Generation and large-scale analysis of expressed sequence tags (EST) from young leaf cDNA library of tea plant (Camellia sinensis (L.) O. Kuntze)

JOURNAL Unpublished (2004)
COMMENT Contact: Chen, Liang
 Laboratory for Germplasm, Breeding and Molecular Biology
 Tea Research Institute, Chinese Academy of Agricultural Sciences
 1 Yundi Road, Hangzhou, Zhejiang 310008, P.R. China
 Tel: 86 571 8665 2835
 Email: cbtri@mail.hz.zj.cn

FEATURES
 source
 1..461
 /organism="Camellia sinensis"
 /mol_type="mRNA"
 /cultiivar="Longjing 43"
 /db_xref="taxon:4442"
 /clone="TL026A02"
 /tissue_type="Young leaf"
 /dev_stage="Growing stage"
 /lab_host="E. coli TGI"
 /clone_lib="Young leaf cDNA library of tea plant (Camellia sinensis)"
 /note="Vector: lambda Triplex2; Site_1: SfiIA and B"

ORIGIN

Query Match 89.6%; Score 22.4; DB 7; Length 461;

Best Local Similarity 95.8%; Pred. No. 6.1e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 455 AAAAAAAAACTTCATCATTTAA 432

RESULT 3
CL378878 484 bp DNA linear GSS 19-AUG-2004

LOCUS
DEFINITION
genomic survey sequence.

ACCESSION
CL378878
KEYWORDS
GSS.

SOURCE
Sus scrofa (pig)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.

REFERENCE
1 (bases 1 to 484)
Rogatcheva, M.B., Meyers, S., He, W., Larkin, D.M., Marron, B.M.,
Beaver, J.B. and Schook, L.B.

TITLE
Piggy-BACing the Human Genome: Constructing a Porcine Physical Map
Through Comparative Genomics

JOURNAL
Unpublished (2004)

COMMENT
Other GSSs: RPCI44_425019.f

Contact: Lawrence B. Schook
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 265 5326
Fax: 217 244 5617

Email: schook@uiuc.edu

Clones are derived from the porcine BAC library RPCI-44
(http://www.bacpac.chori.org/porcine42.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@chori.org).

Clones may be purchased from BACPAC Resources
(http://BACPACresources.chori.org). This work was undertaken as part
of the International Swine Genome Sequencing Consortium by
University of Illinois at Urbana Champaign, USA with funds provided
by grant No. AG2002-34480-11828 from USDA-CSREBS and
AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
Initiative)

Plate: 425 row: O column: 19
Seq primer: SP6
Class: BAC ends.

FEATURES
Source Location/Qualifiers

1..484
/organism="Sus scrofa"
/mol_type="genomic DNA"
/strain="four pigs (bred: 37.5% Yorks Landrace and 25%
Meishan)"
/db_xref="taxon:9823"
/clone="RPCI44_425019"
/sex="male"
/cell_type="blood"
/clone_lib="RPCI-44"
/note="Vector: pTRABAC2; Site_1: EcoRI; Site_2: EcoRI;
porcine male BAC library produced by Pieter de Jong"

ORIGIN
Query Match 89.6%; Score 22.4; DB 10; Length 484;
Best Local Similarity 95.8%; Pred. No. 6.1e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 162 AAAAAAAAACTTCATCATTTAA 185

RESULT 4

EM066980/c 368 bp mRNA linear EST 27-MAY-2005

LOCUS
DEFINITION
EM066980 Nori Satoh unpublished cDNA library, cleaving embryo Clona
intestinalis cDNA clone rcic1101d07 3', mRNA sequence.

ACCESSION
EM066980
KEYWORDS
EST.

SOURCE
Clona intestinalis
Clona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Pleurobranchia; Clonidae; Clona.

REFERENCE
1 (bases 1 to 368)
Satou, Y., Shin-1, T., Kohara, Y. and Satoh, N.
Expressed genes in Clona intestinalis (2002c)

TITLE
Unpublished (2002)

JOURNAL
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

COMMENT
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
Source Location/Qualifiers

1..368
/organism="Clona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rcic1101d07"
/tissue_type="whole body"
/dev_stage="cleaving embryo"
/clone_lib="Nori Satoh unpublished cDNA library, cleaving
embryo"

ORIGIN

Query Match 87.2%; Score 21.8; DB 5; Length 368;
Best Local Similarity 92.0%; Pred. No. 1e+03; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 2;

Qy 1 AAAAAAAAACTTCATCATTTAA 25
Db 146 AAAAAAAAACTTCATCATTTAA 122

RESULT 5
BX158115 566 bp DNA linear GSS 28-JAN-2003

LOCUS
DEFINITION
BX158115 Danio rerio genomic clone DKEX-147118, genomic survey sequence.

ACCESSION
BX158115
KEYWORDS
GSS.

SOURCE
Danio rerio (zebrafish)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

REFERENCE
1 (bases 1 to 566)
Cyriniiformes; Cyprinidae; Danio.

AUTHORS
Humphray, S.J., Huckle, E. and Durham, J.L.
TITLE
Direct Submission

JOURNAL
Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humphray@sanger.ac.uk Unpublished

COMMENT
This sequence was generated from the 7' end of BAC 147118. 147118
is part of the Daniokey BAC library created by R. Plasterk and N.V.
Keygene. Further details:
http://www.sanger.ac.uk/projects/D_rerio/

FEATURES
Source Location/Qualifiers

1..566
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-147118"
/tissue_type="Testis"
/note="vector pindigobAC-536"

ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 566;
 Best Local Similarity 92.0%; Pred. No. 1e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAAC 25
 |||||
 Db 376 AAAAAAAAACTTCATCATTTAAAC 400

RESULT 6
 C2935905/c 976 bp DNA linear GSS 11-AUG-2005
 LOCUS 259827 Tomato EcorI BAC Library Lycopersicon esculentum genomic
 DEFINITION clone SL_EcorI0041A19 3, genomic survey sequence.
 ACCESSION C2935905
 VERSION C2935905.1 GI:72266166
 KEYWORDS GSS.
 SOURCE Lycopersicon esculentum (Solanum lycopersicum)
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 976)
 Mueller, L.A., Buels, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J.,
 Van Eck, J., and Stack, S.
 BAC end sequencing from three Solanum Lycopersicon libraries
 Unpublished (2005)
 Other GSSs: 254981
 Contact: Lukas Mueller
 Tanksley Lab, Dept. of Plant Breeding
 Cornell University
 251 Emerson Hall, Ithaca, NY 14853, USA
 Tel: 607-255-6557
 Fax: 607-255-6683
 Email: sgn-feedback@sgn.cornell.edu
 Plate: 41 row: A column: 19
 Seq primer: SP6
 Class: BAC ends
 High quality sequence start: 50
 High quality sequence stop: 662.
 Location/Qualifiers
 1..976
 /organism="Lycopersicon esculentum"
 /mol_type="genomic DNA"
 /culivar="Heinz 1706"
 /db_xref="taxon:4081"
 /clone="SL_EcorI0041A19"
 /lab_host="E. coli"
 /clone_1fb="Tomato EcorI BAC Library"
 /note="Vector: unk; Site_1: EcorI"

ORIGIN

Query Match 87.2%; Score 21.8; DB 10; Length 976;
 Best Local Similarity 92.0%; Pred. No. 1e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAAC 25
 |||||
 Db 512 AAAAAAAAACTTCATCATTTAAAC 488

RESULT 7
 CV014040/c 537 bp mRNA linear EST 19-AUG-2004
 LOCUS TL012F10 Young leaf cDNA library of tea plant (Camellia sinensis)
 DEFINITION Camellia sinensis cDNA clone TL012F10 5', mRNA sequence.
 ACCESSION CV014040
 VERSION CV014040.1 GI:51453254
 KEYWORDS EST.
 SOURCE Camellia sinensis (black tea)
 ORGANISM Camellia sinensis

REFERENCE
 AUTHORS Chen, L., Zhao, L.P., and Gao, Q.K.
 TITLE Generation and large-scale analysis of expressed sequence tags (EST) from young leaf cDNA library of tea plant (Camellia sinensis (L.) O. Kuntze)
 JOURNAL Unpublished (2004)
 COMMENT Contact: Chen, Liang
 Laboratory for Germplasm, Breeding and Molecular Biology
 Tea Research Institute, Chinese Academy of Agricultural Sciences
 1 Yung Road, Hangzhou, Zhejiang 310008, P.R. China
 Tel: 86 571 8665 2835
 Email: tbrt@mail.hz.zj.cn
 Seq primer: 5'lambda Triplex2 primer CTCGAGATCTGACGAGCT.
 Location/Qualifiers
 1..537
 /organism="Camellia sinensis"
 /mol_type="mRNA"
 /culivar="Longjing 43"
 /db_xref="taxon:4442"
 /clone="TL012F10"
 /issue_type="Young leaf"
 /dev_stage="Growing stage"
 /lab_host="E. coli TG1"
 /clone_1fb="Young leaf cDNA library of tea plant (Camellia sinensis)"
 /note="Vector: lambda Triplex2; Site_1: SfiIA and B"

ORIGIN

Query Match 85.6%; Score 21.4; DB 7; Length 537;
 Best Local Similarity 95.7%; Pred. No. 1.4e+03;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTTCATCATTTAA 24
 |||||
 Db 415 AAAAAAAAACTTCATCATTTAA 393

RESULT 8
 CV014702/c 545 bp mRNA linear EST 19-AUG-2004
 LOCUS TL027B01 Young leaf cDNA library of tea plant (Camellia sinensis)
 DEFINITION Camellia sinensis cDNA clone TL027B01 5', mRNA sequence.
 ACCESSION CV014702
 VERSION CV014702.1 GI:51453916
 KEYWORDS EST.
 SOURCE Camellia sinensis (black tea)
 ORGANISM Camellia sinensis
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; Ericales; Theaceae; Camellia.
 1 (bases 1 to 545)
 Chen, L., Zhao, L.P., and Gao, Q.K.
 Generation and large-scale analysis of expressed sequence tags (EST) from young leaf cDNA library of tea plant (Camellia sinensis (L.) O. Kuntze)
 JOURNAL Unpublished (2004)
 COMMENT Contact: Chen, Liang
 Laboratory for Germplasm, Breeding and Molecular Biology
 Tea Research Institute, Chinese Academy of Agricultural Sciences
 1 Yung Road, Hangzhou, Zhejiang 310008, P.R. China
 Tel: 86 571 8665 2835
 Email: tbrt@mail.hz.zj.cn
 Seq primer: 5'lambda Triplex2 primer CTCGAGATCTGACGAGCT.
 Location/Qualifiers
 1..545
 /organism="Camellia sinensis"
 /mol_type="mRNA"
 /culivar="Longjing 43"
 /db_xref="taxon:4442"
 /clone="TL027B01"

```

/cisue_type="Young leaf"
/dev_stage="Growing stage"
/lab_host="E. coli TG1"
/clone_lib="Young leaf cDNA library of tea plant (Camellia sinensis)"
/note="Vector: lambda Triplex2; Site_1: SfiI and B"

```

Query Match	85.6%	Score 21.4	DB 7	Length 545
Best Local Similarity	95.7%	Pred. No. 1.4e+03		
Matches 22	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	2	AAAAAAAAACCTTCATCATTTTAA	24	
Db	377	AAAAAAAAACCTTCATCATTTTAA	355	

RESULT 9	AV024200/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
	AV024200	234 bp	mus musculus adult C57BL/6J lung	AV024200	AV024200.1	GI:4801192	Mus musculus (house mouse)
		EST 28-AUG-1999	linear				
			mus musculus CDNA clone				

REFERENCE
1 (bases 1 to 234)
Carninci, P., Shibata, K., Ozawa, Y., Komio, H., Itoh, M., Aizawa, K.,

AUTHORS
Achimichi, P., Shibata, K., Ozawa, Y., Kuno, H., Itoh, M., Atzawa, K.,
Ahnira, S., Akiyama, J., Fukuda, Y., Fukunishi, Y., Funayama, T.,
Hara, A., Hayakatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izeawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niiitsuma, H., Oda, H.,
Oswa, K., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE
Journal of
Unpublished (1999)

JOURNAL
Contact: Chie Owa

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel.: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermotranslitzation and thermoinactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp/>) for further details.

```

FEATURES
source
location/Qualifiers
1..234
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="120001H14"
/class_type="lung"
/dev_stage="adult"
/clone_idb="Mus musculus adult C57BL/6J lung"

```

Query Match	84.0%;	Score 21;	DB 1;	Length 234;
Best Local Similarity	100.0%;	Pred. No. 1.9e+03;		
Matches 21; Conservative	0;	Mismatches	0;	Gaps 0;
QY	2	AAAAAAAAACCTTCATCTTAA	22	

Db 230 AAAAAAAAACTTCATCATTTA 210

RESULT	10
LOCUS	BG736368/c
DEFINITION	BG736368 342 bp mRNA linear EST 10-JUN-2001
ACCSSION	RF55603.Y1 Meloidogyne javanica egg pmr1 v6 Chappellet McCarter
VERSION	MG736368
KEYWORDS	Meloidogyne javanica cDNA 5', mRNA sequence.
SOURCE	MG736368.1 GI:14086057
	EST.
	Meloidogyne javanica (root-knot nematode)

ORGANISM	REFERENCE
Meloidogynae javanica	1 (bases 1 to 342)
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogynae.	
Mccarter, J., Clifton, S., Chiappelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Mair, M., Hillier, L., Kuaba, T., Theising, B., Tsagersheivilli, R., Ritter, E., Bennett, J., Franklin, C., Tsagersheivilli, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurr, R., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.	
The Washington Univ. Nematode EST Project, 1999	
Unpublished (1999)	

TITLE
The Washington Univ. Nematode EST Project, 1999

JOURNAL
Unpublished (1999)

COMMENT
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James McCarter (bchiapell@watson.wustl.edu & jmcarter@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
High quality sequence stop: 334.

```

FEATURES
    source
        1.342
            location/Qualifiers
                /organism="Meloidogyne javanica"
                /mol_type="mRNA"
                /db_xref="taxon:6303"
                /dev_stage="enriched for eggs"
                /lab_host="DH10B"
                /clone.lib="Meloidogyne javanica egg pAMP1 v6 Chiappelli
                McCarter"
                /note="Vector: pAMP1 (Gibco) ; The library was constructed
                by Brandi Chiappelli and Dr. James McCarter at Washington
                University, St. Louis. The cDNA was made by using Dynabead
                oligo-dT priming (Dyna)l . PCR based library using a
                modified protocol from the SMART PCR cDNA Synthesis Kit
                from Clontech. Directionally cloned into the XbaI sites of
                pAMP1. Nematodes were provided by Dr. David Bird of North
                Carolina State University."

```

ORIGIN

Query Match	84.0%;	Score 21;	DB 2;	Length 342;
Best Local Similarity	100.0%;	Pred. No. 2e+03;		
Matches 21; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

RESULT 11	
LOCUS	AQ996468/c
DEFINITION	344 bp DNA linear GSS-24-FEB-2000 RPCT-23-355N9.TV RPCT-23 Mus musculus genomic clone RPCT-23-355N9,
ACCESSION	AQ996468
VERSION	AQ996468
KEYWORDS	AQ996468.1 GI:7071565 GSS.

SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 344)
AUTHORS	Zhao,S., Niemann,W., Feldblum,T., Malek,J., Shatnam,S., Aknrcet,B., Levine,M., McGann,S., Tesgaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCI-23-355N9.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html Plate: 355 row: N column: 9 Seq primer: T7 Class: BAC ends.
FEATURES	Location/Qualifiers 1..344 /organism="Mus musculus" /mol_type="genomic DNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="RPCI-23-355N9" /sex="Female" /lab_host="RD10B" /clone_lib="RPCI-23" /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methyase. Site selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
ORIGIN	Query Match 84.0%; Score 21; DB 9; Length 344; Beat Local Similarity 100.0%; Pred. No. 2e+03; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY	1 AAAAAAAAACTTCATCATTT 21
DB	142 AAAAAAAAACTTCATCATTT 122
RESULT 12	
LOCUS	BE447256 421 bp mRNA linear EST 25-JUL-2000
DEFINITION	ut50g01.x1 Soares mouse 3BMS Mus musculus CDNA clone IMAGE:3331344
ACCESSION	BE447256
VERSION	BE447256.1 GI:9446833
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 421)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL: Unpublished (1997)
 COMMENT: Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LBL; contact the
 IMAGE Consortium (info@image.lbl.gov) for further information.
 MGI:1075508.

FEATURES
 source location/Qualifiers
 1..421
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3331344"
 /sex="male"
 /tissue_type="Spleen"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /clone_id="Soares mouse 3MBWS"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCTGAGTGGAGCGCGCGCTGTTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through
 three rounds of normalization, and was constructed by
 Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Query Match 84.0%; Score 21; DB 2; Length 421;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 2 AAAAAAAAACTTCATCATTTA 22
 |||||||||
 Db 29 AAAAAAAAACTTCATCATTTA 49

RESULT 13
 BG583394/c 650 bp mRNA linear EST 11-APR-2001
 LOCUS
 DEFINITION
 mRNA sequence.
 mRNA accession
 BG583394.1 GI:13598458
 ESR.
 MEDICAGO truncatula (barrel medic)
 MEDICAGO truncatula
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosoid I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
 Medicago.
 1 (bases 1 to 650)
 Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S.,
 Town,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
 ESTs from one month old nitrogen-fixing root nodules of Medicago
 truncatula, 2001
 Unpublished (2001)
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borling Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 University of Minnesota name: M584472e TIGR sequence name:
 M581577K More information is available at: <http://www.medicago.org>
 Seq primer: Skmod (CTA GAA CTA gtc gat CC).
 Location/Qualifiers
 1..650
 /organism="Medicago truncatula"
 /mol_type="mRNA"

/cultivar="genotype A17"
 /db_xref="taxon:380"
 /clone="pGVN-73117"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month post_inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /clone_lib="GVN"
 /note="Vector: pluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the uni-ZAP XR vector from Stracagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells."

ORIGIN

Query Match 84.0%; Score 21; DB 2; Length 650;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTT 21
 |||
 Db 431 AAAAAAAAACTTCATCATTT 431

RESULT 14
 LOCUS DN340148 299 bp mRNA linear EST 04-MAR-2005
 DEFINITION LIB5523-014-P1-K1-D2 LIB3523 Canis familiaris cDNA clone
 ACCESSION DN340148
 VERSION DN340148
 KEYWORDS EST
 SOURCE DN340148.1 GI:60512840
 ORGANISM Canis familiaris (dog)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.

FEATURES

REFERENCE 1 (bases 1 to 299)
 AUTHORS Staten,N.R.
 TITLE Direct Submission (Staten,N.R.)
 JOURNAL Unpublished (2005)
 COMMENT Contact: Nick Staten
 Tel: 636 247 6855
 Email: nicholas.r.staten@fizer.com.
 FEATURES
 Source Location/Qualifiers
 1..299
 /organism="Canis familiaris"
 /mol_type="mRNA"
 /accession="Beagle"
 /db_xref="taxon:9615"
 /clone="CAN2089932"
 /tissue_type="brain stem"
 /lab_host="DH10B"
 /clone_lib="LIB3523"
 /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; control dog brain stem"

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 299;
 Best Local Similarity 91.7%; Pred. No. 2.3e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
 |||
 Db 263 AAAAAAAAACTTCATCATTTAA 266

RESULT 15
 BG012600

LOCUS BG012600 308 bp mRNA linear EST 24-JAN-2001
 DEFINITION IL5-GN0239-131200-333-e06 GN0239 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG012600
 VERSION BG012600.1 GI:12461951
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 308)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Ngai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 10737800
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

COMMENT

This sequence was derived from the PARESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL5&t=IL5-GN0239-131200-333-e06&t3=2000-12-13&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 308.

FEATURES

Source Location/Qualifiers
 1..308
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="GN0239"
 /note="Organ: placenta normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 308;
 Best Local Similarity 91.7%; Pred. No. 2.3e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
 |||
 Db 89 AAAAAAAAACTTCATCATTTAA 112

RESULT 16

CR876225/c 373 bp DNA linear GSS 19-NOV-2004
 LOCUS CR876225
 DEFINITION Sus scrofa BES, genomic survey sequence.
 ACCESSION CR876225
 VERSION CR876225.1 GI:55875361
 KEYWORDS GSS; Bac-end sequence BES; Genome Survey Sequence.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suidae; Suidae.

REFERENCE

1 (bases 1 to 373)

AUTHORS Rogel-Gallard, C., Bourdeaux, N., Billault, A., Vaiman, M. and Charbon, P.

TITLE Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements

JOURNAL Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)

PMID 10449899

REFERENCE 2 (bases 1 to 373)

AUTHORS Charbon, P., Iannuccelli, N., Roig, A., Dossat, C., Demars, J., Rogel-Gallard, C., Roy, A., Schibler, L. and Milan, D.

TITLE A physical map of the swine genome

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 373)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

FEATURES - Web : www.genoscope.cns.fr

Source Location/Qualifiers

1..373

/organism="Sus scrofa"

/mol_type="genomic DNA"

/strain="Large White"

/db_xref="taxon:9823"

/clone="b10841G11"

/sex="male"

/cell_type="fibroblast"

/clone_lib="SBAB"

/note="Genoscope sequence ID : IH0AA82BH02FM1"

ORIGIN

Query Match 83.2%; Score 20.8; DB 11; Length 373;

Best Local Similarity 91.7%; Pred. No. 2.3e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAACTCATCTTTAA 24

324 AAAAAAAAACTCATCTAGTAAA 301

RESULT 17 398 bp mRNA linear EST 29-JUN-2004

LOCUS CO349733

DEFINITION DR_AOV_F101.C09 adult ovary full-length (TLL) Danio rerio cDNA, mRNA sequence.

ACCESSION CO349733

VERSION CO349733.1 GI:49431050

KEYWORDS EST.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

AUTHORS Li, Y., Chia, J.M., Bartfai, R., Christofels, A., Yue, G.H., Ke, D., Ho, M.Y., Hill, J.A., Stupka, E. and Orban, L.

TITLE Comparative analysis of the testis and ovary transcriptomes in zebrafish by combining experimental and computational tools

JOURNAL Unpublished (2004)

COMMENT Contact: laezlo@tll.org.sg

REPRODUCTIVE GENOMICS GROUP

REMARKS Lifesciences Laboratory

1 Research Link, The NUS, Singapore 117 604

Tel: +65 6872 7413

Fax: +65 6872 7007

Email: laezlo@tll.org.sg

High quality sequence scop: 398.

FEATURES Location/Qualifiers

Source 1..398

/organism="Danio rerio"

/mol_type="mRNA"

/strain="AB"

/db_xref="taxon:7955"

/sex="female"

AUTHORS Rogel-Gallard, C., Bourdeaux, N., Billault, A., Vaiman, M. and Charbon, P.

TITLE Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements

JOURNAL Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)

PMID 10449899

REFERENCE 2 (bases 1 to 373)

AUTHORS Charbon, P., Iannuccelli, N., Roig, A., Dossat, C., Demars, J., Rogel-Gallard, C., Roy, A., Schibler, L. and Milan, D.

TITLE A physical map of the swine genome

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 373)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

FEATURES - Web : www.genoscope.cns.fr

Source Location/Qualifiers

1..373

/organism="Sus scrofa"

/mol_type="genomic DNA"

/strain="Large White"

/db_xref="taxon:9823"

/clone="b10841G11"

/sex="male"

/cell_type="fibroblast"

/clone_lib="SBAB"

/note="Genoscope sequence ID : IH0AA82BH02FM1"

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 398;

Best Local Similarity 88.0%; Pred. No. 2.3e+03;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy 1 AAAAAAAAACTCATCTTTAA 25

326 AAAAAAAAACTCATCTTTATAC 302

RESULT 18 427 bp mRNA linear EST 25-APR-2001

LOCUS BF654733

DEFINITION 279021 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BF654733

VERSION BF654733.1 GI:11919865

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G., Pettea, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)

PMID 11282978

COMMENT Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68913-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@meat.usda.gov

Single pass sequencing. Bases called and alt. trimmed with phred v0.960904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR PRIMERS

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCACGACGACGACG

Plate: 73 row: C column: 18

Seq primer: ATTGAGTGACCTATAG.

FEATURES Location/Qualifiers

Source 1..427

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/tissue_type="pooled"

/lab_host="DH10B"

/clone_lib="MARC 3BOV"

/note="Vector: PCMV SPOR6; Site 1: NotI; Site 2: SalI; library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 427;
 Best Local Similarity 91.7%; Pred. No. 2.3e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
 |||||
 Db 239 AAAAAAAAACTTCATCATTTAA 262
 |||||

RESULT 19
 AI620206 454 bp mRNA linear EST 21-APR-1999
 LOCUS tus4d07.x1 NCI_CGAP_P128 Homo sapiens cDNA clone IMAGE:2254861 3',
 DEFINITION mRNA sequence.
 ACCESSION AI620206
 VERSION AI620206.1 GI:4629332
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.
 1 (bases 1 to 454)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILMB at:
 www-bio.illn.gov/bbtp/image/image.html
 Seq primer: -400P from Glbco.
 Location/Qualifiers
 1..454
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2254861"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP_P128"
 /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
 with a modified polylinker; Plasmid DNA from the
 normalized library NCI CGAP P128 was prepared, and se
 circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 of 5,000 clones made from the same library (cloneids
 985608-986159, 1101192-1101959, and 1217928-1220615).
 Subtraction by Bento Soares and M. Fatima Bonaudo."

ORIGIN
 Query Match 83.2%; Score 20.8; DB 1; Length 454;
 Best Local Similarity 91.7%; Pred. No. 2.3e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
 |||||
 Db 349 AAAAAACATCTTCATCATTTAA 372
 |||||

RESULT 20
 DR39N7T 516 bp DNA linear GSS 22-NOV-2002
 LOCUS Dario rerio genomic clone DKEX-39N7, genomic survey sequence.
 DEFINITION

ACCESSION AL974669
 VERSION AU974669.1 GI:25182104
 KEYWORDS GSS.
 SOURCE Dario rerio (zebrafish)
 ORGANISM Dario rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 516)
 REFERENCE Humphray, S.V., Huckle, E. and Hunt, S.E.
 JOURNAL Direct Submission
 Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
 humphraysanger.ac.uk Unpublished
 This sequence was generated from the T7 end of BAC 39N7. 39N7 is
 part of the Dariokey BAC library created by R. Plastark and N.V.
 Keygene.
 Further details: http://www.sanger.ac.uk/Projects/D_rerio/.

COMMENT

FEATURES
 source
 Location/Qualifiers
 1..516
 /organism="Dario rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEX-39N7"
 /tissue_type="Testis"
 /note="vector pindigoBAC-536"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 11; Length 516;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
 |||||
 Db 511 AAAAAAACATTCATCATTTAA 488
 |||||

RESULT 21
 AV765288 547 bp mRNA linear EST 18-AUG-2004
 LOCUS AV765288/c
 DEFINITION corniculatus var. japonicus young plants (two-week old) Lotus
 corniculatus var. japonicus cDNA clone MM003906_f_3', mRNA
 sequence.
 ACCESSION AV765288
 VERSION AV765288.1 GI:45348589
 KEYWORDS EST.
 SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
 ORGANISM Lotus corniculatus var. japonicus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
 Lotus.
 1 (bases 1 to 547)
 REFERENCE Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 JOURNAL Characteristics of the Lotus japonicus gene repertoire deduced from
 PUBMED large-scale expressed sequence tag (EST) analysis
 Plant Mol. Biol. 54 (3), 405-414 (2004)
 15284495
 CONTACT: Erika Asamizu
 COMMENT The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 Location/Qualifiers
 1..547
 /organism="Lotus corniculatus var. japonicus"
 /mol_type="mRNA"
 /isolate="Miyakojima MG-20"
 /db_xref="taxon:34305"
 /clone="MM003906.f"
 /tissue_type="whole plant"
 /dev_stage="two-week old plant"
 /clone_lib="Lotus japonicus young plants (two-week old)"

/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; synonym: Lotus japonicus"

Query Match 83.2%; Score 20.8; DB 1; Length 547;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTTAA 24
|||||
Db 354 AAAAAAAAACTTCATCTTTGAA 331

RESULT 22
LOCUS CO911063/c 582 bp mRNA linear EST 16-AUG-2004
DEFINITION BJ03014E05 Capsicum annuum cDNA 5', mRNA sequence.
ACCESSION CO911063
VERSION CO911063.1 GI:51301366
KEYWORDS EST.
SOURCE Capsicum annuum
ORGANISM Capsicum annuum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanaceae; Capsicum.

REFERENCE 1 (bases 1 to 582)
Song, W.Y., Jeon, W.B., Kim, K.S., Lee, H.H., Ko, M.K., Kim, Y.S., Hong, J.C. and Oh, B.J.

TITLE Generation of Hot Pepper (Capsicum annuum) ESTs (Express Sequence Tags) from Red Ripe Fruit (Song, et al.)

JOURNAL Unpublished (2004)

COMMENT Contact: Boung-Jun Oh
Jeonnam Biotechnology Research Center
Namyang Bld. #603, 10-4 Gwangsan-dong, Dong-gu, Gwangju 501-180, Korea

Tel: 82 62 607 2422
Fax: 82 62 607 6203
Email: bjo@biohub.re.kr
Plate: 014 row: E column: 05.

FEATURES
source Location/Qualifiers

1..582
/organism="Capsicum annuum"
/mol_type="mRNA"
/cultiivar="Nokkang"
/db_xref="taxon:4072"
/tissue_type="red pepper fruit pericarp"
/clone_lib="BJ03"
/note="Vector: pBluescript II SK(+):XR; Site_1: EcoRI; Site_2: XhoI; cDNA library was generated from red ripe fruit pericarp using lambda Zap II phage vector. In vivo excision was done with helper phage to generate subclone in pBluescript II SK(+):XR vector."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 582;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTTAA 24
|||||
Db 74 AAAAAAAAACTTCATCTTTAA 51

RESULT 23
LOCUS AG242106/c 586 bp DNA linear GSS 19-JUL-2003
DEFINITION Lotus corniculatus var. japonicus DNA, clone:J1T10b09_not, genomic survey sequence.
ACCESSION AG242106
VERSION AG242106.1 GI:26552943
KEYWORDS GSS.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae; Lotus.

REFERENCE 1
Sato, S., Nakamura, Y. and Tabata, S.
Lotus japonicus TAC End sequences
Published Only in Database (2002)
2 (bases 1 to 586)

REFERENCE
AUTHORS Sato, S.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2002) Shusei Sato, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, 2-6-7
Kazusa-Kametari, Kisarazu, Chiba 292-0818, Japan
(E-mail: ssato@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/, Tel: 81-438-52-3935 (ex. 2336), Fax: 81-438-52-3934)

FEATURES
source Location/Qualifiers

1..586
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/strain="Miyakojima MG-20"
/variety="japonicus"
/db_xref="taxon:34305"
/clone_lib="J1T10b09_not"
/clone_lib="genomic TAC library"
/note="VECTOR: pYTAC7
synonym: Lotus japonicus"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 586;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCTTTAA 24
|||||
Db 473 AAAAAAAAACTTCATCTTTAA 450

RESULT 24
LOCUS BI961316/c 587 bp mRNA linear EST 22-OCT-2001

DEFINITION MONO1_B_H05_b1_A005 Monocytes (MONO1) Equus caballus cDNA, mRNA

ACCESSION BI961316 587 bp
VERSION BI961316.1 GI:16319519
KEYWORDS EST.
SOURCE Equus caballus (horse)
ORGANISM Equus caballus

REFERENCE 1 (bases 1 to 587)
Vandeplass, M.L., Cordonnier-Pratt, M.-M., Sudhan, M.L., Wentzel, V.E., Gingie, A.R., Pratt, L.H. and Moore, J.N.

TITLE An EST database from equine (Equus caballus) monocytes
JOURNAL Unpublished (2001)

COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210

Email: mmp@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions below phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyTmax or T7 sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV
High quality sequence stop: 587
POLYA=No.

FEATURES
source Location/Qualifiers

1..587
/organism="Equus caballus"
/mol_type="mRNA"

/db_xref="taxon:9796"
 /cell_type="isolated peripheral blood monocytes stimulated
 with E. coli lipopolysaccharide"
 /clone_lib="Monocytes (MONO1)"
 /note="Vector: pBluescript SK(-) from lambda ZapII;
 Site 1: XhoI; Site 2: EcoRI. The library was made from
 poly-A RNA in the cloning vector lambda ZapII. Clones to
 be sequenced were prepared by mass excision."

ORIGIN

Query Match 83.2%; Score 20.8; DB 3; Length 587;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTTCATCATTTAA 24
 Db 429 AAAAAAAAACTTCATCATTTCAA 406

RESULT 25

CV014565/c
 LOCUS TL024D05 607 bp mRNA linear EST 19-AUG-2004
 DEFINITION Camellia sinensis cDNA library of tea plant (Camellia sinensis)
 Camellia sinensis cDNA clone TL024D05 5', mRNA sequence.

ACCESSION CV014565
 VERSION CV014565.1 GI:51453779
 KEYWORDS EST.
 SOURCE Camellia sinensis (black tea)
 ORGANISM Camellia sinensis

REFERENCE Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; Ericales; Theaceae; Camellia.
 1 (bases 1 to 607)

AUTHORS Chen, L., Zhao, L. P. and Gao, Q. K.
 TITLE Generation and large-scale analysis of expressed sequence tags
 (EST) from young leaf cDNA library of tea plant (Camellia sinensis
 (L.) O. Kuntze)

JOURNAL

COMMENT Unpublished (2004)
 Contact: Chen, Liang
 Laboratory for Germplasm, Breeding and Molecular Biology
 Tea Research Institute, Chinese Academy of Agricultural Sciences
 1 Yundi Road, Hangzhou, Zhejiang 310008, P.R. China
 Tel: 86 571 8665 2835
 Email: cbrli@mail.hz.zj.cn

Seq primer: 5'-lambda Triplex2 primer CTCGAGATCTGACGAGCT.

FEATURES
 Location/Qualifiers

1..607
 /organism="Camellia sinensis"
 /mol_type="mRNA"
 /cultivar="Longjing 43"
 /db_xref="taxon:4442"
 /clone="TL024D05"
 /tissue_type="Young leaf"
 /dev_stage="Growing stage"
 /lab_host="E. coli TGI"
 /clone_lib="Young leaf cDNA library of tea plant (Camellia
 sinensis)"
 /note="Vector: lambda Triplex2, Site_1: SfiIA and B"

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 607;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTTCATCATTTAA 24
 Db 487 AAAAAAAAACTTCATCATTTAA 464

RESULT 26
 BH765063 643 bp DNA linear GSS 20-MAR-2002
 LOCUS BH765063
 DEFINITION BMBAC53D10SP6_P5U Brugia malayi Genomic Bac Library 3 Brugia

malayi genomic, genomic survey sequence.

BH765063
 VERSION BH765063.1 GI:19562827
 GSS.

KEYWORDS Brugia malayi

SOURCE Brugia malayi

ORGANISM Brugia malayi

REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Brugia.
 1 (bases 1 to 643)

AUTHORS Whitton, C., Daub, J., Quail, M., Hall, N., Foster, J., Ware, J.,
 Ganatra, M., Slack, B., Barrell, B. and Blaxter, M.

A genome sequence survey of the filarial nematode Brugia malayi:
 repeats, gene discovery, and comparative genomics

Mol. Biochem. Parasitol. 137 (2), 215-227 (2004)

15383292

CONTACT: Blaxter ML

Institute of Cell, Animal and Population Biology

University of Edinburgh

Aberforth Labs, King's Buildings, West Mains Road, Edinburgh, EH9

3JT, UK

Tel: +44 131 650 6760

Fax: +44 131 670 5450

Email: mark.blaxter@ac.uk

Sequenced from the Brugia malayi BAC library constructed by Claire

Whitton and Dr Mike Quail. The sequence was generated by The

Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in

collaboration with Mark Blaxter, ICAFP, University of Edinburgh,

Edinburgh, UK.

Seq primer: SP6 (ATTAGTACACTATAG)

Class: BAC ends.

Location/Qualifiers

1..643

/organism="Brugia malayi"

/mol_type="genomic DNA"

/strain="TBS"

/db_xref="taxon:6279"

/sex="Mixed (male and female)"

/tissue_type="whole parasite"

/dev_stage="microfilaria (L1)"

/clone_lib="Brugia malayi Genomic Bac Library 3"

/note="Vector: pBAC3.6; Site 1: BamH I; Brugia malayi

genomic DNA was partially cleaved with Sau3a I and size

fractionated. 7,392 clones were generated with mean insert

size ~48 kbp. The library was constructed by Claire

Whitton, Blaxter Nematode Genetics Lab, University of

Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing

Unit, The Sanger Centre, Cambridge, UK."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 643;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTTCATCATTTAA 24
 Db 398 AAAAAAAAACTTCATCATTTATA 421

RESULT 27
 BB088272/c
 LOCUS BB088272 661 bp mRNA linear EST 31-AUG-2001
 DEFINITION BB088272 RIKEN full-length enriched, 12 days embryo, embryonic body
 between diaphragm region and neck Mus musculus cDNA clone
 9430015G03 3', mRNA sequence.

ACCESSION BB088272
 VERSION BB088272.2 GI:15407718
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 661)

AUTHORS	TITLE
Atakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,T., Komno,H., Koude,M., Koya,S., Matuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.	Riken Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)	
On Jun 21, 2000 this sequence version replaced gi:8654066. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216	
Email: genome-resgsc.riken.jp, URL:http://genome.gsc.riken.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh.M., Komno.H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)	
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsuhira,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.	
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1751-1771 (2000)	
Komno,H., Fukuishi,Y., Shibata,K., Itoh.M., Carninci,P., Sugahara.Y. and Hayashizaki,Y.	
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)	
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,K., Fukuda,S., Hara,A., Itoh.M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.	
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)	
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
Location/Qualifiers	
1..661	
/organism="Mus musculus"	
/mol_type="mRNA"	
/db_xref="taxon:10090"	
/clone="9430015G03"	
/tissue_type="embryonic body between diaphragm region and neck"	
/dev_stage="12 days embryo"	
/lab_host="DH10B"	
/clone_lib="RIKEN full-length enriched, 12 days embryo, embryonic body between diaphragm region and neck"	
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'	
GAGGAGAAGAGATCCACAGACTCTTTTTCCTTTTTTNN 3'] cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAAGATTCGACTGTAAATAAATTATCCCCCCCCCCC	

```

ORIGIN
3'. CDNA was cleaved with XhoI and BamHI."

Query Match 83.2%; Score 20.8; DB 1; Length 661;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAACCTTCATCATTTAA 24
Db AAAAAAAAAACCTTCATGTTTAA 53

RESULT 28
AG314967 676 bp DNA linear GSS 18-DEC-2004
LOCUS AG314967
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-098012.T7, genomic survey
sequence.
ACCESSION AG314967
VERSION AG314967.1 GI:47887924
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1
Aba,K., Noguchi,H., Tagawa,K., Yuzurika,M., Toyoda,A., Kojima,T.,
Rawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiroishi,T.
Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
Genome Res. 14 (12), 2439-2447 (2004)
15574823
2 (bases 1 to 676)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
Direct Submision
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, url:http://ngp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@crc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@crc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
location/Qualifiers
1..676
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-098012.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_id="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 83.2%; Score 20.8; DB 10; Length 676;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 AAAAAAAAAACCTTCATCATTTAA 24

```

Db 233 AAAAAAAAACTTCCTCACTTAA 256

RESULT 29
BI451869
LOCUS
DEFINITION BI451869 687 bp mRNA linear EST 21-AUG-2001
intratradices cDNA clone GI01_A08 3', mRNA sequence.

ACCESSION BI451869
VERSION BI451869.1 GI:15276576
KEYWORDS
SOURCE
ORGANISM
Glomus intratradices
Glomus intratradices
Eukaryota; Fungi; Glomeromycota; Glomeromycetes; Glomerales;
Glomeraceae; Glomus.

REFERENCE
AUTHORS 1 (bases 1 to 687)
Dewbre,G.R., Maldonado-Mendoza,I.E., Bell,C.J. and Harrison,M.J.
ESTs from the extra-radical mycelium of Glomus intratradices during
arbuscular mycorrhizal symbiosis with Daucus carota

JOURNAL
COMMENT Unpublished (2001)
Contact: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Insert Length: 687 Std Error: 0.00
Plate: 01 row: A column: 08
Seq primer: CTCGGAGAGCGGCCCATTTGTGTG.

FEATURES
source
Location/Qualifiers
1..687
/organism="Glomus intratradices"
/mol_type="mRNA"
/db_xref="taxon:4876"
/clone="GI01_A08"
/issue_type="Extra-radical mycelium"
/note="Vector: Lambda Triplex2; Extra-radical mycelium
from Glomus intratradices grown in a two-compartment
petri-plate system in association with Daucus carota
transformed roots. M medium containing 35 micromolar
potassium phosphate was added to the extra-radical
mycelium 12 hours prior to harvest. cDNA was prepared
from total RNA using the Clontech Smart cDNA synthesis
system. The cDNA was directionally ligated into Lambda
Triplex2 vector (Clontech) and packaged using Gigapack
Gold packaging extracts. Plasmids (pTriplex2) containing
cDNA inserts were obtained from the recombinant lambda
Triplex2 phage via Cre-lox mediated conversion in BM25.8
cells (Clontech)."

ORIGIN
Query Match 83.2%; Score 20.8; DB 3; Length 687;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 102 AAAAAAAAAATCATCATTTAA 125

RESULT 30
CE502891/c
LOCUS
DEFINITION CE502891 696 bp DNA linear GSS 28-SEP-2003
tigr-gss-dog-17000310740927 Dog Library Canis familiaris genomic,
genomic survey sequence.

ACCESSION CE502891
VERSION CE502891.1 GI:36819672
KEYWORDS
SOURCE
ORGANISM
Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS 1 (bases 1 to 696)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)

JOURNAL
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkne@tigr.org
Class: Shotgun.

FEATURES
source
Location/Qualifiers
1..696
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BctXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 83.2%; Score 20.8; DB 10; Length 696;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 63 AAAAAAAAAATTCACCATTTAA 40

RESULT 31
BH567772/c
LOCUS
DEFINITION BH567772 739 bp DNA linear GSS 14-DEC-2001
BOHCG92TP BOHC Brassica oleracea genomic clone BOHCG92, genomic
survey sequence.

ACCESSION BH567772
VERSION BH567772.1 GI:17819612
KEYWORDS
SOURCE
ORGANISM
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
AUTHORS 1 (bases 1 to 739)
Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
Utterback,T.R., Mortman,J.R., White,O.R. and Tompa,C.D.
Whole genome shotgun sequencing of Brassica oleracea and its
application to gene discovery and annotation in Arabidopsis
Genome Res. 15 (4), 487-495 (2005)

JOURNAL
COMMENT Other GSSs: BOHCG92TR
15805490
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.

FEATURES
source
Location/Qualifiers
1..739
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 739;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
 |||||
 713 AAAAAAAAAATCTCATATTAA 690

Db

RESULT 32
 LOCUS CK362049/c 740 bp mRNA linear EST 23-DEC-2003
 DEFINITION AGENCOURT_1717202 NIH_ZGC_4 Danio rerio cDNA clone IMAGE:7086265
 ACCESSION CK362049
 VERSION CK362049
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 740)
 NIH-MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Daniela S. Gerhard, Ph.D.
 COMMENT Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgaps-remail.nih.gov
 Tissue Procurement: John Ngai, Univ of CA, Berkeley
 cDNA Library Preparation: Dr. Sumio Sugano
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LHM14914 row: j column: 23
 High quality sequence stop: 541.
 Location/Qualifiers
 1..740
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:7086265"
 /lab_host="DH10B Tona"
 /clone_1lb="NIH_ZGC_4"
 /note="Organ: brain/CNS; Vector: pME18S-FL3; Site_1: DraIII; Site_2: DraIII"

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 740;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
 |||||
 520 AAAAAAAAAAGCTTCATGATTAA 497

Db

RESULT 33
 LOCUS CO253171/c 761 bp mRNA linear EST 23-JUN-2004
 DEFINITION WS00818_B21.1_O22 WS-X-N-A-9 Picea glauca cDNA clone WS00818_O22
 3', mRNA sequence.
 ACCESSION CO253171

VERSION CO253171.1 GI:49134688
 KEYWORDS EST.
 SOURCE Picea glauca (white spruce)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
 1 (bases 1 to 761)
 Ralph,S., Kolosova,N., Oddy,C., Cooper,D., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Babkaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Friedman,M.F., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Matra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.
 The spruce transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries
 unpublished (2005)
 JOURNAL Contact: Joerg Bohlmann
 Genome BC forest genomics program
 University of British Columbia
 Michael Smith Laboratories, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3
 Tel: 1-604-822-0282
 Fax: 1-604-822-2114
 Email: bohlmann@msl.ubc.ca
 Plate: WS00818 row: 0 column: 22
 High quality sequence stop: 761
 POLYA=Yes.
 Location/Qualifiers
 1..761
 /organism="Picea glauca"
 /mol_type="mRNA"
 /cultiivar="Pg-29"
 /db_xref="taxon:3330"
 /clone="WS00818_O22"
 /sex="Hermaphrodite"
 /tissue_type="Barkly season xylem harvested June 15th, mid season xylem harvested July 10th and late season xylem harvested August 17th"
 /lab_host="E. coli DH10B cells"
 /clone_1lb="WS-X-N-A-9"
 /note="Organ: Outer xylem from 25 year old trees harvested at Kalamalka Research Station in Vernon, British Columbia in 2001; Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5' end of cDNA); Site_2: XhoI (3' end of cDNA); mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldi M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

ORIGIN

Query Match 83.2%; Score 20.8; DB 7; Length 761;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
 |||||
 50 AAAAAAAAAACCTTCATGATTAA 27

Db

RESULT 34
 LOCUS DN527189 764 bp mRNA linear EST 11-MAR-2005
 DEFINITION 1272019 MARC 7BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION DN527189
 VERSION DN527189.1 GI:60972421

KEYWORDS
SOURCE EST
ORGANISM Bos taurus (cow)

REFERENCE
AUTHORS Smith, T.P.L., Roberts, A.J., Echeverkamp, S.E., Chitko-McKown, C.G., Wray, J.E. and Keeler, J.W.
TITLE A second set of bovine ESTs from pooled-tissue normalized libraries
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smtlth@mail.marc.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_aln option. Vector identified with cross_match v0.990329.
 Plate: RUK8047 row: B column: 22
 Seq primer: GTATACGACCTCATATAGGG.

FEATURES
source
 Location/Qualifiers
 1..764
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /rname="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 780V"
 /note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI; Library made with RNA pooled from multiple tissues including ovary, hindbrain, uterus, and day-30 whole embryos."

ORIGIN

Query Match 83.2%; Score 20.8; DB 8; Length 764;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 AAAAAAAAAACCTCATCTTTAA 24
 615 AAAAAAAAAACCTCATCTTTAA 638

RESULT 35
CC573268
LOCUS CH240_449J24.TARBA13P2 CHORI-240 Bos taurus genomic clone
DEFINITION CH240_449J24, genomic survey sequence.
ACCESSION CC573268
VERSION CC573268.1 GI:31912771
KEYWORDS GSS.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
REFERENCE
AUTHORS Holte, R., Stort, J., Yang, G., Barber, S., Smalhus, D., Prabhu, A.-L., Tsai, M., Cloutier, A., Lee, D., Gilm, N., Olson, T., Mayo, M., Butterfield, Y., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Chiu, R., Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S., Schein, J., Marra, M., de Jong, P., Keeler, J.W. and Kappes, S.M.
TITLE Bovine BAC End Sequences from Library CHORI-240, PLATES 399 to 478
JOURNAL Unpublished (2003)
COMMENT Other GSSs: CH240_449J24.T7
 Contact: Rob Holte
 Sequencing
 The British Columbia Cancer Agency Genome Science Centre
 600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4G6
 Tel: 604-877-6085
 Fax: 604-877-6276

Email: rholt@bccsc.ca
 Clones are derived from the bovine BAC library CHORI-240
 (<http://www.chori.org/bacpac/bovine240.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources
 (<http://www.chori.org/bacpac/ordering/information.htm>). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada.
 Plate: 449 row: J column: 24
 Seq primer: SP6
 Class: BAC ends.

FEATURES
source
 Location/Qualifiers
 1..773
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="bred: Hereford"
 /db_xref="taxon:9913"
 /clone="CH240_449J24"
 /sex="Male"
 /cell_type="Blood"
 /note="Vector: pTARBA1.3; Site 1: MboI; Site 2: MboI; Hereford bull 11 Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 773;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2 AAAAAAAAAACCTCATCTTTAAC 25
 52 AAAAAAAAAACCTCATCTTTAAC 75

RESULT 36
AG316741
LOCUS AG316741 794 bp DNA linear GSS 18-DEC-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-101E16.T7, genomic survey sequence.
ACCESSION AG316741
VERSION AG316741.1 GI:47889698
KEYWORDS GSS.
SOURCE Mus musculus molossinus (Japanese wild mouse)
ORGANISM Mus musculus molossinus
REFERENCE
AUTHORS Abe, K., Noguchi, H., Tagawa, K., Yuzurika, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaki, K. and Shirotshi, T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)
COMMENT 2 (bases 1 to 794)
 Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 Direct Submission
 Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenho-chou, Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:<http://hgp.gsc.riken.go.jp/>, Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the mouse BAC library MSMG01. For BAC library availability, please contact Kunya Abe (abe@rtc.riken.jp).
 Tsukuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyatai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp

PRIMERS
Sequencing : T7
LIBRARY : pBACe3.6
Vector : EcoRI
R.Site 1 : EcoRI
R.Site 2 : EcoRI

FEATURES
Location/Qualifiers
1..794
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSM901-101E6.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSM901 Mouse Male BAC Library"

ORIGIN
Query Match 83.2%; Score 20.8; DB 10; Length 794;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTTCATCTTAA 24
|||
334 AAAAAAAAACTTCCTCTTAA 357

Db 334 AAAAAAAAACTTCCTCTTAA 357

RESULT 37
DU088441 837 bp DNA linear GSS 12-AUG-2005
DU088441/c 36073 Tomato HindIII BAC Library Lycopersicon esculentum genomic
LOCUS DU088441
DEFINITION DU088441 GI:72524582
VERSION DU088441.1 GI:72524582
KEYWORDS
SOURCE
ORGANISM Lycopersicon esculentum (Solanum lycopersicum)
Lycopersicon esculentum
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 837)
Weller, L.A., Buel, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J.,
Van Eck, J. and Stack, S.
BAC end sequencing from three Solanum lycopersicon libraries
Unpublished (2005)
Other GSSes: 36072
Contact: Lukas Mueller
Tanksley Lab, Dept. of Plant Breeding
Cornell University
251 Emerson Hall, Ithaca, NY 14853, USA
Tel: 607-255-6557
Fax: 607-255-6683
Email: segn-feedback@cornell.edu
Insert Length: 104031 Std Error: 0.00
Plate: 111 row: H column: 13
Seq primer: T7
Class: BAC ends
High quality sequence start: 45
High quality sequence stop: 702.
Location/Qualifiers
1..837
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/cultivar="Heinz 1706"
/db_xref="taxon:4081"
/clone="LE_HBa011H13"
/lab_host="E. coli"
/clone_lib="Tomato HindIII BAC Library"
/note="Vector: pBeloBAC11, Site_1: HindIII"

ORIGIN
Query Match 83.2%; Score 20.8; DB 10; Length 837;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTTCATCTTAA 24
|||
821 AAAAAAAAACTTCCTTAA 798

Db 821 AAAAAAAAACTTCCTTAA 798

RESULT 38
DR882998 839 bp mRNA linear EST 01-AUG-2005
LOCUS DR882998
DEFINITION UGI CABR1506 rev NIH XGC tropSp11 Xenopus tropicalis cDNA clone
IMAGE:7761380 3', mRNA sequence.
DR882998
DR882998.1 GI:71572250
EST.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 839)
Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,
Brokstein, P. and Lindquist, R.A.
DOB Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: UGI_CABR1506.fwd
Contact: Lindquist, E.A., Richardson, P.
DOB Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Robert M. Grainger
CDNA Library Preparation: Bruce Blumberg Laboratory, University of
California, Irvine
DNA Sequencing: DOB Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LML:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of
the UGI Clone id and the direction of sequencing. The suffix '.rev'
indicates a reverse sequencing read of the insert. It does not
necessarily reflect the orientation of the insert.
Poly-A: Based upon the presence of a run of 14 or more T residues
at the beginning of the sequence, this clone was polyadenylated.
The resulting Poly-T sequence has been removed.
Plate: CABK 0013 row: d column: 18
High quality sequence stop: 773
POLYA=Yes.
Location/Qualifiers
1..839
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/strain="N6 (Nigerian 6th generation inbred)"
/db_xref="taxon:8364"
/clone="IMAGE:7761380"
/tissue_type="Spleen"
/dev_stage="Adult"
/lab_host="ElectroMAX DH10B T1 Phage Resistant cells"
/clone_lib="NIH XGC tropSp11"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The
library was prepared from total RNA by oligo-dT priming
(5'-ACTAGTCGCGCCCTCAGGCTCAGGCTTCTTTTCTTTT-3') and
Stratascript reverse transcriptase. After ligation of
EcoRI adaptors (5'-AATTCGACGAGG-3') followed by kinasing
adapters and by XhoI digestion, the cDNA was size selected
by chromatography on sepharose CL-2B columns and fractions
containing cDNAs larger than 1000 bp were ligated into
EcoRI/XhoI-digested pCS107. Reference for library
construction: Current Genetics 4, 635-644. Library
developmental and Cell Biology, University of California,
Irvine."

ORIGIN

```

REFERENCE
AUTHORS      1 (bases 1 to 891)
              Mireva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,
              Rafter,E., Martin,J., Mylle,T., Dante,M., Waterston,R.H.,
              Clifton,S.W. and Wilson,R.
JOURNAL      Genome survey sequences from the rat parasitic nematode
COMMENT      Strongyloides ratii
              Unpublished (2005)
              Washington University in St. Louis
              Contact: Mireva M
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel.: 314 286 1800
              Fax: 314 286 1810
              Email: nematode@watson.wustl.edu
              Genomic DNA was provided by Fiona Thompson
              (F.Thompson@bristol.ac.uk) and Mark Viney
              (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol,
              UK.

FEATURES
source       Class: shotgun.
              Location/Qualifiers
              1..891
                /organism="Strongyloides ratii"
                /mol_type="genomic DNA"
                /strain="Isotemale line ED321 heterogonic"
                /db_xref="taxon:34506"
                /dev_stage="infective larval stage (L3)"
                /lab_host="G310"
                /clone_lib="Strongyloides ratii whole genome shotgun
                library (SRA:SS 004)"
                /note="vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
                Strongyloides ratii genomic DNA was randomly sheared,
                end-repaired and size fractioned to enrich for 2-4 kb
                fragments. Genomic DNA was provided by Fiona Thompson
                (F.Thompson@bristol.ac.uk) and Mark Viney
                (Mark.Viney@bristol.ac.uk) at the University of Bristol,
                Bristol, UK. Sequencing by Washington University Genome
                Sequencing Center, St. Louis, MO."

ORIGIN
Query Match      83.2%; Score 20.8; DB 10; Length 891;
Best Local Similarity 91.7%; Pred. No. 2,4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0.

QY
1 AAAAAAAAAACCTTCATCATTTAA 24
  |||||
  |||||
Db      780 AAAAAAGATACCTTCATCATTTAA 757

RESULT 41
LOCUS      AE539375/c
DEFINITION AE539375 891 bp DNA linear GSS 14-NOV-2000
ACCESSION  AE539375
VERSION    AE539375
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica
ORGANISM   Entamoeba histolytica
REFERENCE  1 (bases 1 to 891)
AUTHORS    Loftus,B., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
JOURNAL    HMI:IMSS sheared DNA library
COMMENT    Unpublished (2000)
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjlloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library

```

Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 18
 High quality sequence stop: 775.
 Location/Qualifiers

FEATURES

source

1.891
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMS8"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica sheared DNA"
 /note="Vector: pPOS1; Site 1: Bet 1; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 Kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 891;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCTTTAA 24
 |||||
 Db 818 AAAAAAAAAACCTTCATCTTTAA 795

RESULT 42

CF579446/c 894 bp mRNA linear EST 24-SEP-2003
 DEFINITION AGENCOURT 10675679 updated NIH_MGC_137 Mus musculus cDNA clone
 IMAGE:6434746 5', mRNA sequence.

CF579446
 CF579446.1 GI:35192445
 EST.
 Mus musculus (house mouse)
 Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.

1 (bases 1 to 894)
 NIH-MGC http://mgs.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgsdbs-remail.nih.gov
 Tissue Procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000)
 cDNA Library Preparation: Catherine Lee, Endocrine Pancreas
 Consortium

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 plate: IRB50 row: g column: 11
 High quality sequence stop: 571.
 Location/Qualifiers

FEATURES

source

1.894
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6434746"

/lab_host="DH10B"
 /clone_lib="NIH_MGC_137"
 /note="Organ: pancreas; Vector: pSPORT1; Site 1: SalI;
 Site 2: NotI; Library consists of a pool of clones
 rearrayed from the following libraries: Melton normalized
 mixed mouse pancreas 1 NI-MMS1, Amplified Melton mouse
 1beta 1 Misi-A, and Kaestner right wt. Clones rearrayed in
 the laboratory of K. Kaestner (University of
 Pennsylvania). Note: this is a NIH_MGC library."

ORIGIN

Query Match 83.2%; Score 20.8; DB 6; Length 894;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCTTTAA 24
 |||||
 Db 890 AAAAAAAAAACCTTCATCTTTAA 867

RESULT 43

CZ993563 938 bp DNA linear GSS 11-AUG-2005
 DEFINITION 180709 Tomato Mbol BAC library Lycopersicon esculentum genomic
 clone SL_Mbol10062612 3, genomic survey sequence.

ACCESSION CZ993563
 VERSION CZ993563.1 GI:72345208
 KEYWORDS GSS.
 SOURCE Lycopersicon esculentum (Solanum lycopersicum)
 ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; lamids; Solanales; Solanales; Solanales; Lycopersicon.

1 (bases 1 to 938)
 Mueller, L.A., Bueis, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J.,
 Van Eck, J. and Stack, S.
 BAC end sequencing from three Solanum lycopersicon libraries
 Unpublished (2005)
 Other GSSs: 180710
 Contact: Lukas Mueller
 Tanksley Lab, Dept. of Plant Breeding
 Cornell University
 251 Emerson Hall, Ithaca, NY 14853, USA
 Tel: 607-255-6557
 Fax: 607-255-6683
 Email: sgn-feedback@sgn.cornell.edu
 Plate: 62 row: G column: 12
 Seq primer: SP6
 Class: BAC ends
 High quality sequence start: 7
 High quality sequence stop: 343.
 Location/Qualifiers

FEATURES

source

1.938
 /organism="Lycopersicon esculentum"
 /mol_type="genomic DNA"
 /cultivar="Heinz 1706"
 /db_xref="taxon:4081"
 /clone="SL_Mbol10062612"
 /lab_host="E. coli"
 /clone_lib="Tomato Mbol BAC library"
 /note="Vector: pBelobAC11; Site 1: Mbol"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 938;
 Best Local Similarity 91.7%; Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCTTTAA 24
 |||||
 Db 514 AAAAAAAAAACCTTCATCTTTAA 491

RESULT 44

BU167817/c
 LOCUS BU167817 981 bp mRNA linear EST 04-SEP-2002
 DEFINITION AGENCOURT 8076667 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6089011
 5', mRNA sequence.
 ACCESSION BU167817
 VERSION BU167817.1 GI:22681801
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.
 1 (bases 1 to 981)
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DP
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LCM2328 row: b column: 20
 High quality sequence stop: 659.
 Location/Qualifiers
 1..981
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6089011"
 /tissue_type="melanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_112"
 /note="Organ: skin; Vector: pOT7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN
 Query Match 83.2% Score 20.8; DB 5; Length 981;
 Best Local Similarity 91.7% Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTTCATCATTTAA 24
 Db 896 AAAAAAAAACTTCATTTATTTAA 873

RESULT 45
 BH157390 1014 bp DNA linear GSS 24-SEP-2001
 LOCUS BH157390
 DEFINITION ENTOMX57F Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, genomic survey sequence.
 ACCESSION BH157390
 VERSION BH157390.1 GI:15730828
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 1014)
 REFERENCE Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HM1:IMSS sheared DNA library (2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics

The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 81
 High quality sequence stop: 344.
 Location/Qualifiers
 1..1014
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt; Site_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."

ORIGIN
 Query Match 83.2% Score 20.8; DB 9; Length 1014;
 Best Local Similarity 91.7% Pred. No. 2.4e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTTCATCATTTAA 24
 Db 671 AAAAAAAAACTTTATTTATTTAA 694

RESULT 46
 CNS00GLK 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS CNS00GLK/c
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
 BACR33018 of RCT-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL071937
 VERSION AL071937.1 GI:4952121
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 REFERENCE Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mawmser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named Rpt-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Source

1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR33018"
/note="end : TET3"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 1101;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 64 AAAAAAAAACTTATCATTTAA 41

RESULT 47
CC257387 1161 bp DNA linear GSS 13-MAY-2003
LOCUS CH261-166P19.RM1.1 CH261 Gallus gallus genomic clone CH261-166P19,
DEFINITION genomic survey sequence.
ACCESSION CC257387
VERSION CC257387.1 GI:30597085
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1161)
Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
Warren,W., Graves,T., Mardis,E. and Wilson,R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: RM1 TAGGACTCACTATAGGAGCA
Class: BAC ends
High quality sequence start: 23
High quality sequence stop: 693.
Location/Qualifiers
1..1161

FEATURES

Source

1..1161
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-166P19"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - for library and clone
ordering information: <http://www.choxi.org/bacpac>"

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 1161;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 746 AAAAAAAAACTTCATCATTTCAA 769

RESULT 48
CL039316/c 1177 bp DNA linear GSS 31-DEC-2003
LOCUS CH216-47L13.RM1.1 CH216 Xenopus tropicalis genomic clone
DEFINITION CH216-47L13, genomic survey sequence.
ACCESSION CL039316
VERSION CL039316
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 1177)
Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM1 TAGGACTCACTATAGGAGCA
Class: BAC ends
High quality sequence start: 134
High quality sequence stop: 405.
Location/Qualifiers
1..1177

FEATURES
Source
1..1177
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-47L13"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

ORIGIN

Query Match 83.2%; Score 20.8; DB 10; Length 1177;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 145 AAAAAAAAACTTCATCATTTAA 122

RESULT 49

CC289898 1211 bp DNA linear GSS 13-MAY-2003
LOCUS CH261-34H12.RM1.1 CH261 Gallus gallus genomic clone CH261-34H12,
DEFINITION genomic survey sequence.
ACCESSION CC289898
VERSION CC289898.1 GI:30661339
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1211)
Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
Warren,W., Graves,T., Mardis,E. and Wilson,R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu

SOURCE

REFERENCE

AUTHORS Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
Warren,W., Graves,T., Mardis,E. and Wilson,R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu

TITLE

JOURNAL

COMMENT

Insert Length: 182000 Std Error: 0.00
Seq primer: RM1 TACGACTCACTATAGGAGACA

Class: BAC ends
High quality sequence start: 36
High quality sequence stop: 755.
Location/Qualifiers

FEATURES
source

1. 1211
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-34H12"
/sex="female"
/cell_line="UCDD001, indred 256"
/clone_id="CH261"
/note="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: <http://www.choxi.org/bacpac>"

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 1211;
Best Local Similarity 91.7%; Pred. No. 2.4e+03;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 1033 AAAAAAAAACTTCATCATTTATA 1056

RESULT 50

CC230460

CH261-183K24_Sp6.1 CH261 Gallus gallus genomic clone CH261-183K24,

LOCUS

genomic survey sequence.

CC230460

CC230460.1 GI:30557123

GSS.

Gallus gallus (chicken)

SOURCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 1457)

Kremetzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,

Warren, W., Graves, T., Mardis, E. and Wilson, R.

Gallus gallus BAC End Reads

Unpublished (2003)

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@wustl.edu

Insert Length: 182000 Std Error: 0.00

Seq primer: Sp6 ATTAGGTGACACTATAG

Class: BAC ends

High quality sequence start: 256

High quality sequence stop: 920.

Location/Qualifiers

1. 1457

/organism="Gallus gallus"

/mol_type="genomic DNA"

/strain="Red Jungle Fowl"

/db_xref="taxon:9031"

/clone="CH261-183K24"

/sex="female"

/cell_line="UCDD001, indred 256"

/clone_id="CH261"

/note="Vector: PTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;

CH261 Female Chicken library - For library and clone

ordering information: <http://www.choxi.org/bacpac>"

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAA 24
Db 973 AAAAAAAAACTTCATCATTTCAA 996

Search completed: December 14, 2005, 07:35:37
Job time: 1761.1 secs

ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 1457;
Best Local Similarity 91.7%; Pred. No. 2.5e+03;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:43:33 ; Search time 180.2 Seconds
(without alignments)
68.002 Million cell updates/sec

Title: US-10-681-773-9

Perfect score: 25

Sequence: 1 aaaaaaaaaaccatcattcaaac 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4161359 seqs, 245077644 residues

Total number of hits satisfying chosen parameters: 8322718

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database: Published Applications NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB_seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB_seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB_seq.*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB_seq.*
5: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PUB_seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB_seq.*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB_seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB_seq.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB_seq.*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB_seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.4	81.6	121001	7	US-11-117-187-190 Sequence 190, App
2	19.2	76.8	1206	6	US-10-750-185-24968 Sequence 24968, A
3	19.2	76.8	1658	6	US-10-750-185-50500 Sequence 50500, A
4	19.2	76.8	1767	6	US-10-750-185-50141 Sequence 50141, A
5	19.2	76.8	2319	6	US-10-750-185-54132 Sequence 54132, A
6	18.6	74.4	774	6	US-10-750-185-50424 Sequence 60424, A
7	18.6	74.4	816	6	US-10-750-185-40972 Sequence 40972, A
8	18.6	74.4	1317	6	US-10-750-185-46604 Sequence 46604, A
9	18.6	74.4	1403	6	US-10-750-185-42013 Sequence 42013, A
10	18.6	74.4	1409	6	US-10-750-185-34073 Sequence 34073, A
11	18.6	74.4	1989	6	US-10-750-185-34070 Sequence 34070, A
12	18.6	74.4	2401	7	US-11-041-776-81 Sequence 81, App
13	18.6	74.4	2761	7	US-11-194-246-381 Sequence 381, App
14	18.6	74.4	156297	7	US-11-121-086-65 Sequence 65, App
15	18.2	72.8	943	6	US-10-750-185-54492 Sequence 54492, A
16	18.2	72.8	1189	6	US-10-750-185-37376 Sequence 37376, A
17	18.2	72.8	3224	8	US-11-112-944-23 Sequence 23, App
18	17.8	71.2	1063	6	US-10-750-185-59346 Sequence 49346, A
19	17.8	71.2	1272	6	US-10-750-185-59757 Sequence 39757, A
20	17.8	71.2	1328	6	US-10-750-185-61956 Sequence 61956, A
21	17.8	71.2	387780	6	US-10-995-561-13259 Sequence 13259, A
22	17.6	70.4	201	6	US-10-995-561-15189 Sequence 15189, A
23	17.6	70.4	600	6	US-10-750-185-21634 Sequence 21634, A

24	17.6	70.4	661	6	US-10-750-185-50164 Sequence 50164, A
25	17.6	70.4	865	6	US-10-750-185-51209 Sequence 51209, A
26	17.6	70.4	1039	6	US-10-750-185-49134 Sequence 49134, A
27	17.6	70.4	1106	6	US-10-750-185-61005 Sequence 61005, A
28	17.6	70.4	1203	6	US-10-525-710-9 Sequence 9, App
29	17.6	70.4	1283	6	US-10-750-185-10646 Sequence 40646, A
30	17.6	70.4	1309	6	US-10-750-185-61694 Sequence 61694, A
31	17.6	70.4	1314	6	US-10-750-185-47958 Sequence 47958, A
32	17.6	70.4	1348	6	US-10-750-185-43373 Sequence 43373, A
33	17.6	70.4	1572	6	US-10-131-826A-373 Sequence 373, App
34	17.6	70.4	1677	6	US-10-750-185-61533 Sequence 61533, A
35	17.6	70.4	1719	6	US-10-750-185-54932 Sequence 24932, A
36	17.6	70.4	1816	6	US-10-750-185-58392 Sequence 58392, A
37	17.6	70.4	1853	6	US-10-750-185-38831 Sequence 38831, A
38	17.6	70.4	1920	6	US-10-750-185-62360 Sequence 62360, A
39	17.6	70.4	2238	6	US-10-793-626-1741 Sequence 1741, App
40	17.6	70.4	2300	6	US-10-750-185-63271 Sequence 63271, A
41	17.6	70.4	2478	6	US-10-750-185-52236 Sequence 52236, A
42	17.6	70.4	2570	6	US-10-750-185-59624 Sequence 59624, A
43	17.6	70.4	3073	6	US-10-793-626-3581 Sequence 3581, App
44	17.6	70.4	3308	6	US-10-750-185-31042 Sequence 31042, A
45	17.6	70.4	4519	6	US-10-750-185-25830 Sequence 25830, A
46	17.6	70.4	5174	6	US-10-821-234-187 Sequence 187, App
47	17.6	70.4	23959	6	US-10-995-561-13475 Sequence 13475, A
48	17.6	70.4	33042	6	US-10-995-561-13340 Sequence 13340, A
49	17.6	70.4	51870	6	US-10-995-561-13199 Sequence 13199, A
50	17.6	70.4	159497	7	US-11-112-908-61 Sequence 61, App
51	17.6	70.4	163317	7	US-11-117-187-212 Sequence 212, App
52	17.6	70.4	170837	7	US-11-121-086-97 Sequence 97, App
53	17.6	70.4	170993	7	US-11-121-086-35 Sequence 35, App
54	17.6	70.4	180654	7	US-11-121-086-58 Sequence 58, App
55	17.6	70.4	319608	7	US-11-145-703-1 Sequence 1, App
56	17.4	69.6	804	6	US-10-750-185-26945 Sequence 26945, A
57	17.4	69.6	187745	7	US-11-121-086-83 Sequence 83, App
58	17.2	68.8	201	6	US-10-995-561-75652 Sequence 75652, A
59	17.2	68.8	819	6	US-10-750-185-41091 Sequence 41091, A
60	17.2	68.8	1139	6	US-10-750-185-48039 Sequence 48039, A
61	17.2	68.8	1280	6	US-10-750-185-44763 Sequence 44763, A
62	17.2	68.8	1323	6	US-10-750-185-55305 Sequence 55305, A
63	17.2	68.8	1773	6	US-10-750-185-52745 Sequence 52745, A
64	17.2	68.8	1823	6	US-10-750-185-47493 Sequence 47493, A
65	17.2	68.8	2579	6	US-10-750-185-32543 Sequence 32543, A
66	17.2	68.8	3299	6	US-10-750-185-36561 Sequence 36561, A
67	17.2	68.8	3465	6	US-10-750-185-42573 Sequence 42573, A
68	17.2	68.8	3572	6	US-10-793-626-3988 Sequence 3988, App
69	17.2	68.8	3750	6	US-10-955-054A-129 Sequence 129, App
70	17.2	68.8	43943	6	US-10-995-561-13466 Sequence 13466, A
71	17.2	68.8	154548	7	US-11-121-086-33 Sequence 33, App
72	17	68.0	201	6	US-10-995-561-46275 Sequence 46275, A
73	17	68.0	264	6	US-10-467-657-1807 Sequence 1807, App
74	17	68.0	649	6	US-10-750-185-63975 Sequence 63975, A
75	17	68.0	725	6	US-10-750-185-50396 Sequence 50396, A
76	17	68.0	942	6	US-10-750-185-29438 Sequence 29438, A
77	17	68.0	953	6	US-10-750-185-60943 Sequence 60943, A
78	17	68.0	1113	6	US-10-750-185-37814 Sequence 37814, A
79	17	68.0	1128	6	US-10-750-185-30160 Sequence 30160, A
80	17	68.0	1187	6	US-10-750-185-37548 Sequence 37548, A
81	17	68.0	1292	7	US-11-058-817A-7 Sequence 7, App
82	17	68.0	1299	6	US-10-750-185-77952 Sequence 27952, A
83	17	68.0	1426	6	US-10-750-185-48414 Sequence 48414, A
84	17	68.0	1564	6	US-10-750-185-47845 Sequence 28745, A
85	17	68.0	1613	6	US-10-750-185-77466 Sequence 77466, A
86	17	68.0	1650	6	US-10-750-185-50725 Sequence 50725, A
87	17	68.0	1850	6	US-10-750-185-63232 Sequence 63232, A
88	17	68.0	1874	6	US-10-750-185-37767 Sequence 37767, A
89	17	68.0	1881	6	US-10-750-185-52504 Sequence 52504, A
90	17	68.0	2388	6	US-10-750-185-64332 Sequence 64332, A
91	17	68.0	2489	6	US-10-750-185-61943 Sequence 61943, A
92	17	68.0	2565	6	US-10-750-185-50905 Sequence 50905, A
93	17	68.0	3272	6	US-10-750-185-37398 Sequence 37398, A
94	17	68.0	3223	6	US-10-750-185-46677 Sequence 46677, A
95	17	68.0	3455	6	US-10-750-185-25466 Sequence 25466, A
96	17	68.0	3720	6	US-10-750-185-55687 Sequence 55687, A

```

97 17 68.0 6659 6 US-10-750-185-26060
98 17 68.0 98862 7 US-11-121-086-76
99 17 68.0 110711 6 US-10-995-561-13264
100 17 68.0 169495 7 US-11-121-086-61
101 17 68.0 172147 7 US-11-112-908-22
102 17 68.0 175416 7 US-11-121-086-43
103 17 68.0 188682 7 US-11-112-908-33
104 17 68.0 199130 6 US-10-995-561-13243
105 17 68.0 222094 6 US-10-995-561-13244
106 17 68.0 415117 6 US-10-995-561-13274
107 17 68.0 645179 6 US-10-995-561-13293
108 17 68.0 645179 6 US-10-995-561-13293
109 16.8 67.2 201 6 US-10-995-561-47523
110 16.8 67.2 201 6 US-10-995-561-48012
111 16.8 67.2 600 6 US-10-750-185-21219
112 16.8 67.2 663 6 US-10-793-626-1863
113 16.8 67.2 939 6 US-10-793-626-1209
114 16.8 67.2 988 6 US-10-750-185-16870
115 16.8 67.2 1342 6 US-10-750-185-59025
116 16.8 67.2 1930 6 US-10-750-185-37997
117 16.8 67.2 1993 6 US-10-750-185-43425
118 16.8 67.2 2743 6 US-10-821-234-267
119 16.8 67.2 3076 6 US-10-793-626-4302
120 16.8 67.2 3714 6 US-10-793-626-3775
121 16.8 67.2 4114 6 US-10-793-626-4184
122 16.8 67.2 218821 7 US-11-121-086-31
123 16.8 66.4 201 6 US-10-995-561-14784
124 16.8 66.4 201 6 US-10-995-561-49779
125 16.8 66.4 201 6 US-10-995-561-49780
126 16.8 66.4 201 6 US-10-995-561-51636
127 16.8 66.4 201 6 US-10-995-561-53063
128 16.8 66.4 201 6 US-10-995-561-53269
129 16.8 66.4 201 6 US-10-995-561-53270
130 16.8 66.4 201 6 US-10-995-561-79134
131 16.8 66.4 201 6 US-10-995-561-79135
132 16.8 66.4 357 6 US-10-467-657-395
133 16.8 66.4 598 6 US-10-750-185-724
134 16.8 66.4 598 6 US-10-750-185-1203
135 16.8 66.4 598 6 US-10-750-185-1005
136 16.8 66.4 729 6 US-10-750-185-57780
137 16.8 66.4 827 6 US-10-750-185-26194
138 16.8 66.4 863 6 US-10-750-185-40046
139 16.8 66.4 939 6 US-10-750-185-62200
140 16.8 66.4 1053 6 US-10-750-185-32372
141 16.8 66.4 1183 6 US-10-750-185-46128
142 16.8 66.4 1325 6 US-10-750-185-56371
143 16.8 66.4 1366 6 US-10-750-185-36334
144 16.8 66.4 1423 6 US-10-750-185-63291
145 16.8 66.4 1516 6 US-10-750-185-61010
146 16.8 66.4 1537 6 US-10-750-185-58763
147 16.8 66.4 1554 6 US-10-750-185-50036
148 16.8 66.4 1584 6 US-10-750-185-55484
149 16.8 66.4 1784 6 US-10-750-185-18799
150 16.8 66.4 1810 6 US-10-750-185-51565

```

ALIGNMENTS

```

RESULT 1
US-11-117-187-190
; Sequence 190, Application US/11117187
; Publication No. US2005026660A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPELHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219

```

```

; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 190
; LENGTH: 121001
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (2128)..(120379)
; OTHER INFORMATION: N = A, C, G, or T/U
US-11-117-187-190

```

```

Query Match 81.6%; Score 20.4; DB 7; Length 121001;
Best Local Similarity 95.5%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 AAAAAAAAACTTCATCATTTA 22
Db 115019 AAAAAAAAACTTCATCATTTCA 115040

```

```

RESULT 2
US-10-750-185-24968
; Sequence 24968, Application US/10750185
; Publication No. US2005026603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24968
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Bovine 1986880207888
US-10-750-185-24968

```

```

Query Match 76.8%; Score 19.2; DB 6; Length 1206;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 AAAAAAAAACTTCATCATTTAA 24
Db 964 AAAAAAAAACTTCATCATTTAA 987

```

```

RESULT 3
US-10-750-185-50500/c
; Sequence 50500, Application US/10750185
; Publication No. US2005026603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31

```

```

; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50500
; LENGTH: 1658
; TYPE: DNA
; ORGANISM: Bovine 19866880654604
US-10-750-185-50500

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 6; Length 1658;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
Db 821 AATATAAACTTCATCATCTTGAA 798

RESULT 4
US-10-750-185-50141
; Sequence 50141, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50141
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Bovine 19866880701039
US-10-750-185-50141

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 6; Length 1767;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
Db 473 AAAAAAAAACTTCATCATTGATA 496

RESULT 5
US-10-750-185-54132/c
; Sequence 54132, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54132
; LENGTH: 816
```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54132
; LENGTH: 2319
; TYPE: DNA
; ORGANISM: Bovine 19866881332429
US-10-750-185-54132

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 6; Length 2319;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
Db 761 AAAAAAAAACTTAATCATTTAAA 738

RESULT 6
US-10-750-185-60424/c
; Sequence 60424, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60424
; LENGTH: 774
; TYPE: DNA
; ORGANISM: Bovine 19866881742396
US-10-750-185-60424

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 6; Length 774;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAAC 25
Db 89 AAAAAAAAACTTTTATTATTAAAC 65

RESULT 7
US-10-750-185-40972
; Sequence 40972, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40972
; LENGTH: 816
```

```

; TYPE: DNA
; ORGANISM: Bovine 19866881088362
US-10-750-185-40972

```

```

Query Match 74.4%; Score 18.6; DB 6; Length 816;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 196 AAAAAAAAAACCTTCATCATTTAGAAC 220

```

```

RESULT 8
US-10-750-185-46604
; Sequence 46604, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46604
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Bovine 19866881559710
US-10-750-185-46604

```

```

Query Match 74.4%; Score 18.6; DB 6; Length 1317;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 405 AAAAAAAAAACCTTCATCATTTACAC 429

```

```

RESULT 9
US-10-750-185-42013
; Sequence 42013, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42013
; LENGTH: 1403
; TYPE: DNA
; ORGANISM: Bovine 19866881136350
US-10-750-185-42013

```

```

Query Match 74.4%; Score 18.6; DB 6; Length 1403;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 434 AAAAAAAAAACCTTCATCATTTAGAAC 458

```

```

RESULT 10
US-10-750-185-34073
; Sequence 34073, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34073
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Bovine 19866881760657
US-10-750-185-34073

```

```

Query Match 74.4%; Score 18.6; DB 6; Length 1409;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAAACCTTCATCATTTAAAC 25
Db 92 AAAAAAAAAACCTTCATCATTTTGAC 116

```

```

RESULT 11
US-10-750-185-34070
; Sequence 34070, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34070
; LENGTH: 1989
; TYPE: DNA
; ORGANISM: Bovine 19866881343407
US-10-750-185-34070

```

```

Query Match 74.4%; Score 18.6; DB 6; Length 1989;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;

```


Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25

Db 1891 AAAAAAAAACTTCATTTTGAC 1915

RESULT 12

US-11-041-776-81
; Sequence 81, Application US/11041776
; Publication No. US20050272057A1
; GENERAL INFORMATION:
; APPLICANT: ABRAMSEN, MITCHELL
; APPLICANT: FREIJE, WADIAH
; TITLE OF INVENTION: SMALL SEGMENTS OF DNA DETERMINE ANIMAL IDENTITY AND
; TITLE OF INVENTION: SOURCE
; FILE REFERENCE: 34579-97951
; CURRENT APPLICATION NUMBER: US/11/041,776
; PRIOR FILING DATE: 2005-01-24
; PRIOR APPLICATION NUMBER: 60/538,791
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: 60/539,728
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 81
; LENGTH: 2401
; TYPE: DNA
; ORGANISM: Sus scrofa
US-11-041-776-81

Query Match

74.4%; Score 18.6; DB 7; Length 2401;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25

Db 1611 AAAAAAAAACTTCCTTTTAAAC 1635

RESULT 13

US-11-194-246-381
; Sequence 381, Application US/11194246
; Publication No. US20050272089A1
; GENERAL INFORMATION:
; APPLICANT: Mott, John
; APPLICANT: Trepod, Catherine
; APPLICANT: Arvidson, Steffan
; TITLE OF INVENTION: CRITICAL GENES AND POLYPEPTIDES OF HAEMOPHILUS INFLUENZAE AND MET
; TITLE OF INVENTION: USE
; FILE REFERENCE: 00592.US1 (M&R 268, 05920101)
; CURRENT APPLICATION NUMBER: US/11/194,246
; PRIOR FILING DATE: 2005-08-01
; PRIOR APPLICATION NUMBER: US/10/274,586
; PRIOR FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US 60/345,438
; PRIOR FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 621
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 381
; LENGTH: 2761
; TYPE: DNA
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: Sequence of the rps9 and rpl13 genes and flanking regions.
US-11-194-246-381

Query Match

74.4%; Score 18.6; DB 7; Length 2761;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25

Db 2307 AAAAAAAAACTTCATCATTTAAAC 2331

RESULT 14

US-11-121-086-65
; Sequence 65, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 65
; LENGTH: 156297
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-65

Query Match

74.4%; Score 18.6; DB 7; Length 156297;
Best Local Similarity 84.0%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTAAAC 25

Db 53816 AACAAAAAATCTCATTTATGTACNC 53840

RESULT 15

US-10-750-185-54492
; Sequence 54492, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54492
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-54492

Query Match

72.8%; Score 18.2; DB 6; Length 943;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTTCATCATTTAAAC 24

Db 599 ATAAAAAATCTCATGATTTAAAC 621

RESULT 16

US-10-750-185-27376
; Sequence 27376, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:

```

; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27376
; LENGTH: 1189
; TYPE: DNA
; ORGANISM: Bovine 19866881074045
US-10-750-185-27376

Query Match      72.8%; Score 18.2; DB 6; Length 1189;
Best Local Similarity 87.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 23
Db 405 AAAAAAAAAATTCATCATTTCA 427

RESULT 17
US-11-112-944-23/C
; Sequence 23, Application US/11112944
; Publication No. US20050244872A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; TITLE OF INVENTION: Breast Cancer Gene Expression Biomarkers
; FILE REFERENCE: 05-325-US
; CURRENT APPLICATION NUMBER: US/11/112,944
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/566,757
; PRIOR FILING DATE: 2004-04-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 23
; LENGTH: 3224
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-944-23

Query Match      72.8%; Score 18.2; DB 8; Length 3224;
Best Local Similarity 87.0%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAAACCTTCATCATTTAA 24
Db 3102 AAAAAAAAAACCATCATCATCTTAA 3080

RESULT 18
US-10-750-185-49346/C
; Sequence 49346, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
```

```

; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49346
; LENGTH: 1063
; TYPE: DNA
; ORGANISM: Bovine 19866880624730
US-10-750-185-49346

Query Match      71.2%; Score 17.8; DB 6; Length 1063;
Best Local Similarity 90.5%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAAACCTTCATCATTTA 22
Db 1035 AAAAAAAAAACCTCATATTTCA 1015

RESULT 19
US-10-750-185-39757
; Sequence 39757, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39757
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Bovine 19866880943580
US-10-750-185-39757

Query Match      71.2%; Score 17.8; DB 6; Length 1272;
Best Local Similarity 90.5%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTT 21
Db 1018 AAAAAAAAAACCTGATCATTT 1038

RESULT 20
US-10-750-185-61956
; Sequence 61956, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
```

PRIOR FILING DATE: 2002-12-31
 NUMBER OF SEQ ID NOS: 64922
 SOFTWARE: PatentIN version 3.1
 SEQ ID NO 61956
 LENGTH: 1328
 TYPE: DNA
 ORGANISM: Bovine 19866881022604
 US-10-750-185-61956

Query Match 71.2%; Score 17.8; DB 6; Length 1328;
 Best Local Similarity 90.5%; Pred. No. 4.2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTT 21
 DB 398 AAAAAAAAACTTCATAATTT 418

RESULT 21
 US-10-995-561-13259/C
 Sequence 13259, Application US/10995561
 Publication No. US20050272054A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michele et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 FILE REFERENCE: CLO01559
 CURRENT APPLICATION NUMBER: US/10/995,561
 CURRENT FILING DATE: 2004-11-24
 NUMBER OF SEQ ID NOS: 85702
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13259
 LENGTH: 387780
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (1)...(387780)
 OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
 US-10-995-561-13259

Query Match 71.2%; Score 17.8; DB 6; Length 387780;
 Best Local Similarity 90.5%; Pred. No. 5.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAAAAAAACTTCATCATTTAA 23
 DB 150806 ACAACACCTTCATCATTTAA 150786

RESULT 22
 US-10-995-561-15189
 Sequence 15189, Application US/10995561
 Publication No. US20050272054A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michele et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 FILE REFERENCE: CLO01559
 CURRENT APPLICATION NUMBER: US/10/995,561
 CURRENT FILING DATE: 2004-11-24
 NUMBER OF SEQ ID NOS: 85702
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 15189
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-995-561-15189

Query Match 70.4%; Score 17.6; DB 6; Length 201;
 Best Local Similarity 83.3%; Pred. No. 4.2e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTTCATCATTTAA 24
 DB 93 AAAAAAAAACTTCAGAAATGAAA 116

RESULT 23
 US-10-750-185-21634/C
 Sequence 21634, Application US/10750185
 Publication No. US20050260603A1
 GENERAL INFORMATION:
 APPLICANT: MMT GENOMICS, INC.
 APPLICANT: DENISE, Sue K.
 APPLICANT: KERR, Richard
 APPLICANT: ROSENFELD, David
 APPLICANT: HOLM, Tom
 APPLICANT: BATES, Stephen
 APPLICANT: FANTIN, Dennis
 TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 FILE REFERENCE: MM1100-2
 CURRENT APPLICATION NUMBER: US/10/750,185
 CURRENT FILING DATE: 2003-12-31
 PRIOR APPLICATION NUMBER: US 60/437,482
 PRIOR FILING DATE: 2002-12-31
 NUMBER OF SEQ ID NOS: 64922
 SOFTWARE: PatentIN version 3.1
 SEQ ID NO 21634
 LENGTH: 600
 TYPE: DNA
 ORGANISM: Bovine MMBT01612
 US-10-750-185-21634

Query Match 70.4%; Score 17.6; DB 6; Length 600;
 Best Local Similarity 83.3%; Pred. No. 4.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
 DB 221 AAAAAACAAAGCTTCAGAAATTTAA 198

RESULT 24
 US-10-750-185-50164/C
 Sequence 50164, Application US/10750185
 Publication No. US20050260603A1
 GENERAL INFORMATION:
 APPLICANT: MMT GENOMICS, INC.
 APPLICANT: DENISE, Sue K.
 APPLICANT: KERR, Richard
 APPLICANT: ROSENFELD, David
 APPLICANT: HOLM, Tom
 APPLICANT: BATES, Stephen
 APPLICANT: FANTIN, Dennis
 TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 FILE REFERENCE: MM1100-2
 CURRENT APPLICATION NUMBER: US/10/750,185
 CURRENT FILING DATE: 2003-12-31
 PRIOR APPLICATION NUMBER: US 60/437,482
 PRIOR FILING DATE: 2002-12-31
 NUMBER OF SEQ ID NOS: 64922
 SOFTWARE: PatentIN version 3.1
 SEQ ID NO 50164
 LENGTH: 661
 TYPE: DNA
 ORGANISM: Bovine 19866882083820
 US-10-750-185-50164

Query Match 70.4%; Score 17.6; DB 6; Length 661;
 Best Local Similarity 83.3%; Pred. No. 4.7e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24

Db 68 AAAATAATCCTGATCATTTAAA 45

RESULT 25

US-10-750-185-51209
; Sequence 51209, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51209
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Bovine 19866881036126
US-10-750-185-51209

Query Match 70.4%; Score 17.6; DB 6; Length 865;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAAA 24

Db 310 AAAATAAAACGTTTATCATTTAAA 333

RESULT 26

US-10-750-185-49134
; Sequence 49134, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49134
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Bovine 19866880936155
US-10-750-185-49134

Query Match 70.4%; Score 17.6; DB 6; Length 1039;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAAA 24

Db 778 AAAAAAAAAACCTTCATTTAAA 801

RESULT 27

US-10-750-185-61005/c
; Sequence 61005, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61005
; LENGTH: 1106
; TYPE: DNA
; ORGANISM: Bovine 19866881864257
US-10-750-185-61005

Query Match 70.4%; Score 17.6; DB 6; Length 1106;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAAA 24

Db 331 AAAACAAAGCTTCAGATTTTAAA 308

RESULT 28

US-10-525-710-9/c
; Sequence 9, Application US/10525710
; Publication No. US20050260721A1
; GENERAL INFORMATION:
; APPLICANT: Kroger, Burkhard
; APPLICANT: Zeider, Oskar
; APPLICANT: Kolprogge, Corinna
; APPLICANT: Schroder, Hartwig
; APPLICANT: Hafner, Stefan
; TITLE OF INVENTION: Method for Zymotic Production of Fine Chemicals Containing
; FILE REFERENCE: 13111-00006-US
; CURRENT APPLICATION NUMBER: US/10/525,710
; CURRENT FILING DATE: 2005-02-24
; PRIOR APPLICATION NUMBER: PCT/EP 2003/009453
; PRIOR FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: DE 102 39 082.7
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9
; LENGTH: 1203
; TYPE: DNA
; ORGANISM: Bacillus stearothermophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) .. (1203)
US-10-525-710-9

Query Match 70.4%; Score 17.6; DB 6; Length 1203;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAAAAAAAAACCTTCATCATTTAAAC 25

Db 70 AAAATAAACCTTCCTTTCAAAAC 47

```
RESULT 29
US-10-750-185-40646/c
; Sequence 40646, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40646
; LENGTH: 1283
; TYPE: DNA
; ORGANISM: Bovine 1986681538120
US-10-750-185-40646

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 6; Length 1283;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 1067 AAGAAAAACCTTATATTTTAA 1044

RESULT 30
US-10-750-185-61694
; Sequence 61694, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61694
; LENGTH: 1309
; TYPE: DNA
; ORGANISM: Bovine 19866881425805
US-10-750-185-61694

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 6; Length 1309;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 363 AAAAAAACCCACCATATTTTAA 386

RESULT 31
```

```
US-10-750-185-47958
; Sequence 47958, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47958
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Bovine 19866881122710
US-10-750-185-47958

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 6; Length 1314;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 806 AAAAAAACCTTATATTAATTA 829

RESULT 32
US-10-750-185-43373/c
; Sequence 43373, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43373
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Bovine 19866881935088
US-10-750-185-43373

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 6; Length 1348;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 914 AAAAAAGACTTCTCATTTAA 891

RESULT 33
US-10-131-826A-373/c
; Sequence 373, Application US/10131826A
; Publication No. US20050245730A1
```

```
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Denoyere, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerltzen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gunney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ TITLE OF INVENTION: ACIDS ENCODING THE SAME
/ FILE REFERENCE: P330R1C128
/ CURRENT APPLICATION NUMBER: US/10/131,826A
/ CURRENT FILING DATE: 2002-04-24
/ PRIOR APPLICATION NUMBER: 60/049911
/ PRIOR FILING DATE: 1997-06-18
/ PRIOR APPLICATION NUMBER: 60/056974
/ PRIOR FILING DATE: 1997-08-26
/ PRIOR APPLICATION NUMBER: 60/059113
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059115
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059117
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059122
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059184
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059263
/ PRIOR FILING DATE: 1997-09-18
/ PRIOR APPLICATION NUMBER: 60/059352
/ PRIOR FILING DATE: 1997-09-19
/ PRIOR APPLICATION NUMBER: 60/059588
/ PRIOR FILING DATE: 1997-09-19
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 373
/ LENGTH: 1572
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-10-131-826A-373

Query Match      70.4%; Score 17.6; DB 6; Length 1572;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1  AAAAAAAAAACCTTCATCATTTAA 24
Db      1215  AGAAAAAACCCACCTCATTTAA 1192
```

RESULT 34

```
US-10-750-185-61533/c
/ Sequence 61533, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
```

```
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 61533
/ LENGTH: 1677
/ TYPE: DNA
/ ORGANISM: Bovine 19866880669242
US-10-750-185-61533
```

```
Query Match      70.4%; Score 17.6; DB 6; Length 1677;
Best Local Similarity 83.3%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1  AAAAAAAAAACCTTCATCATTTAA 24
Db      927  AAAAAAAAACTTAATAATAATA 904
```

```
RESULT 35
US-10-750-185-24932/c
/ Sequence 24932, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
/ PRIOR APPLICATION NUMBER: US 60/437,482
/ PRIOR FILING DATE: 2002-12-31
/ NUMBER OF SEQ ID NOS: 64922
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 24932
/ LENGTH: 1719
/ TYPE: DNA
/ ORGANISM: Bovine 19866881003671
US-10-750-185-24932
```

```
Query Match      70.4%; Score 17.6; DB 6; Length 1719;
Best Local Similarity 83.3%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1  AAAAAAAAAACCTTCATCATTTAA 24
Db      1285  AATTAAACCTTCATCATTTGTA 1262
```

```
RESULT 36
US-10-750-185-58392/c
/ Sequence 58392, Application US/10750185
/ Publication No. US20050260603A1
/ GENERAL INFORMATION:
/ APPLICANT: MMI GENOMICS, INC.
/ APPLICANT: DENISE, Sue K.
/ APPLICANT: KERR, Richard
/ APPLICANT: ROSENFELD, David
/ APPLICANT: HOLM, Tom
/ APPLICANT: BATES, Stephen
/ APPLICANT: FANTIN, Dennis
/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
/ FILE REFERENCE: MM11100-2
/ CURRENT APPLICATION NUMBER: US/10/750,185
/ CURRENT FILING DATE: 2003-12-31
```

```

; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58392
; LENGTH: 1816
; TYPE: DNA
; ORGANISM: Bovine 19866880532336
US-10-750-185-58392

Query Match          70.4%; Score 17.6; DB 6; Length 1816;
Best Local Similarity 83.3%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
    |||||
Db 1244 AAAAAAAAACTTCCTTTACA 1221

RESULT 37
US-10-750-185-38831/c
; Sequence 38831, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38831
; LENGTH: 1853
; TYPE: DNA
; ORGANISM: Bovine 19866880817583
US-10-750-185-38831

Query Match          70.4%; Score 17.6; DB 6; Length 1853;
Best Local Similarity 83.3%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
    |||||
Db 1501 AATAAATACTTCATCATTTACA 1478

RESULT 38
US-10-750-185-62360
; Sequence 62360, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
US-10-750-185-62360
```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62360
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bovine 1986688118596
US-10-750-185-62360

Query Match          70.4%; Score 17.6; DB 6; Length 1920;
Best Local Similarity 83.3%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
    |||||
Db 854 AAAAAAAAACTTCCTTGAAA 877

RESULT 39
US-10-793-626-1741/c
; Sequence 1741, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1741
; LENGTH: 2238
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-1741

Query Match          70.4%; Score 17.6; DB 6; Length 2238;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTCATCATTTAA 24
    |||||
Db 2066 AATAAATACTTCATCATTTAA 2043

RESULT 40
US-10-750-185-63271/c
; Sequence 63271, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63271
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Bovine 19866880374239
US-10-750-185-63271
```

Query Match 70.4%; Score 17.6; DB 6; Length 2300;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 109 AAAACAACATCATCATTTAAA 86

RESULT 41
US-10-750-185-52236/c
; Sequence 52236, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52236
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Bovine 19866880500300
US-10-750-185-52236

Query Match 70.4%; Score 17.6; DB 6; Length 2478;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 1552 AAAATAAATCTTCAACATTTCAA 1529

RESULT 42
US-10-750-185-59624/c
; Sequence 59624, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59624
; LENGTH: 2570
; TYPE: DNA
; ORGANISM: Bovine 19866880937949
US-10-750-185-59624

Query Match 70.4%; Score 17.6; DB 6; Length 2570;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 405 ATAAAAAATCTTCAATTTTAAA 382

RESULT 43
US-10-793-626-3581/c
; Sequence 3581, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3581
; LENGTH: 3073
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-3581

Query Match 70.4%; Score 17.6; DB 6; Length 3073;
Best Local Similarity 83.3%; Pred. No. 5.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 1632 AATAAATAAGCTTCATCATTAATA 1609

RESULT 44
US-10-750-185-31042
; Sequence 31042, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31042
; LENGTH: 3308
; TYPE: DNA
; ORGANISM: Bovine 198668809391353
US-10-750-185-31042

Query Match 70.4%; Score 17.6; DB 6; Length 3308;
Best Local Similarity 83.3%; Pred. No. 5.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAAAAAAAAACCTTCATCATTTAA 24
Db 1279 AAAAAAAAAATCTTTTATTTTAAA 1302


```
RESULT 45
US-10-750-185-25830/c
; Sequence 25830, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25830
; LENGTH: 4519
; TYPE: DNA
; ORGANISM: Bovine 19866880691515
US-10-750-185-25830

Query Match      70.4%; Score 17.6; DB 6; Length 4519;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTTAA 24
Db 2821 AAAAAAACTATCATCATTTATA 2798

RESULT 46
US-10-821-234-187
; Sequence 187, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmati, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 187
; LENGTH: 5174
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-187

Query Match      70.4%; Score 17.6; DB 6; Length 5174;
Best Local Similarity 83.3%; Pred. No. 5.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTTAA 24
Db 1890 AAAAAAAATCTTAAAGCATTTTAA 1913

RESULT 47
US-10-995-561-13475/c
; Sequence 13475, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
```

```
APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CI001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13475
; LENGTH: 29959
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13475

Query Match      70.4%; Score 17.6; DB 6; Length 29959;
Best Local Similarity 83.3%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTTCATCATTTTAAAC 25
Db 24462 AAAAAAACTACAACAATTAAC 24439

RESULT 48
US-10-995-561-13340/c
; Sequence 13340, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CI001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13340
; LENGTH: 33042
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13340

Query Match      70.4%; Score 17.6; DB 6; Length 33042;
Best Local Similarity 83.3%; Pred. No. 6.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTCATCATTTTAA 24
Db 13682 AAAAAAACTACTTATTTCAA 13659

RESULT 49
US-10-995-561-13199
; Sequence 13199, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CI001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13199
; LENGTH: 151870
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: (1)...(151870)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13199

```

```

Query Match      70.4%: Score 17.6; DB 6; Length 151870;
Best Local Similarity 83.3%: Pred. No. 6.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAACCTTCATCATTTTAA 24
        |||||
Db      58337 AAAAAAAAAACCTTCATAATGAAA 58360

```

```

RESULT 50
US-11-112-908-61
; Sequence 61, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; PRIORITY FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIORITY FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIORITY FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIORITY FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIORITY FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 61
; LENGTH: 159497
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-61

```

```

Query Match      70.4%: Score 17.6; DB 7; Length 159497;
Best Local Similarity 83.3%: Pred. No. 6.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 AAAAAAAAAACCTTCATCATTTTAA 24
        |||||
Db      8105 ATAATATAAATTCATCATTTTAA 8128

```

```

Search completed: December 14, 2005, 11:40:55
Job time : 184.2 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:07:18 ; Search time 861.8 Seconds
(without alignments)
1648.975 Million cell updates/sec

Title: US-10-681-773-10

Perfect score: 25
Sequence: 1 aaaaaaactaaagcttgatcttc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl:*
1: gb_ha:*
2: gb_in:*
3: gb_env:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_srs:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	28145	8	AC124305
2	25	100.0	142745	8	AC124305
3	25	100.0	179810	8	AC124305
4	23.4	93.6	147404	8	AC004738
5	22.4	89.6	628	10	BV456584
6	22.4	89.6	724	10	BV456584
7	22.4	89.6	1681	15	AK221711
8	22.4	89.6	1639	15	AV059899
9	22.4	89.6	59793	15	AB025634
10	21.8	87.2	137278	9	AC104103
11	21.8	87.2	162223	8	AC145988
12	21.4	85.6	88311	14	AL358954
13	21.4	85.6	115335	8	AC093811
14	21.4	85.6	134089	8	AC093811
15	21.4	85.6	145722	14	AC015503
16	21.4	85.6	160936	14	AC073494
17	21.4	85.6	166019	14	AC015648
18	21.4	85.6	196721	8	AC019179

19	20.8	83.2	18525	2	AY821561
20	20.8	83.2	84322	15	AP004509
21	20.8	83.2	110000	15	CR382124_07
22	20.8	83.2	157897	14	AC152800
23	20.8	83.2	187909	14	CR936342
24	20.8	83.2	201275	9	AL603652
25	20.8	83.2	214295	9	AC137947
26	20.8	83.2	217253	14	AC156538
27	20.8	83.2	218875	14	CR931978
28	20.8	83.2	227741	9	AC126071
29	20.8	83.2	230880	14	AC158901
30	20.8	83.2	236253	14	AC161390
31	20.8	83.2	246490	14	AC103058
32	20.8	83.2	262198	14	AC106703
33	20.4	81.6	110000	1	AE017243_1
34	20.4	81.6	110000	1	AE017243_01
35	20.4	81.6	110000	1	AE017332_2
36	20.2	80.8	553	10	BV354105
37	20.2	80.8	608	10	BV327179
38	20.2	80.8	617	2	AF148494
39	20.2	80.8	627	6	AR504707
40	20.2	80.8	654	10	BV267418
41	20.2	80.8	654	10	BV267418
42	20.2	80.8	1560	2	AF043416
43	20.2	80.8	1594	2	AF043416
44	20.2	80.8	1594	2	AF043416
45	20.2	80.8	2154	2	AY434012
46	20.2	80.8	2154	2	AY434012
47	20.2	80.8	2254	2	AY434015
48	20.2	80.8	2261	2	AF043417
49	20.2	80.8	2261	2	AF043417
50	20.2	80.8	4254	6	CQ576746
51	20.2	80.8	4254	6	CQ576746
52	20.2	80.8	6746	6	CQ575117
53	20.2	80.8	254321	14	AC116005
54	20.2	80.8	61020	6	AX251546
55	20.2	80.8	67847	14	AC017241
56	20.2	80.8	70453	14	AC101556
57	20.2	80.8	77275	8	AL353653
58	20.2	80.8	79635	14	BX080806
59	20.2	80.8	86133	14	AC153681
60	20.2	80.8	88899	14	AP003841
61	20.2	80.8	89435	14	AC142019
62	20.2	80.8	95567	14	AC091232
63	20.2	80.8	98476	15	AP003706
64	20.2	80.8	110000	1	CP000025
65	20.2	80.8	110000	15	AP008212_150
66	20.2	80.8	110000	15	AP008212_189
67	20.2	80.8	110000	15	AP008207_088
68	20.2	80.8	110000	15	AP008207_088
69	20.2	80.8	110000	15	AP008207_226
70	20.2	80.8	110000	15	AP008207_274
71	20.2	80.8	110000	15	AP008212_274
72	20.2	80.8	110000	15	AP008213_033
73	20.2	80.8	119525	15	AP006366
74	20.2	80.8	123292	15	AC090870
75	20.2	80.8	127437	8	AC105312
76	20.2	80.8	129402	15	AP003341
77	20.2	80.8	129402	8	HSJ709121
78	20.2	80.8	132427	9	AL354652
79	20.2	80.8	137118	14	AC165218
80	20.2	80.8	138021	15	AC090820
81	20.2	80.8	141678	15	AP004664
82	20.2	80.8	146081	15	AP001366
83	20.2	80.8	146551	15	AP004621
84	20.2	80.8	146536	15	AP005610
85	20.2	80.8	148570	9	AC091514
86	20.2	80.8	151991	15	AC091122
87	20.2	80.8	154298	9	AL631773
88	20.2	80.8	154378	15	AP001383
89	20.2	80.8	157273	9	BX293558
90	20.2	80.8	158081	9	AC110043
91	20.2	80.8	158312	9	AC154408

AY821561 Manduca s
AP004509 Lotus cor
Continuation (8 of
AC152800 Bos tauru
CR936342 Dantio rer
AL603652 Mouse DNA
AC137947 Mus muscu
AC156538 Bos tauru
CR931978 Dantio rer
AC126071 Rattus no
AC158901 Mus muscu
AC161390 Bos tauru
AC103058 Rattus no
AC106703 Rattus no
Continuation (2 of
Continuation (2 of
Continuation (3 of
BV354105 S230P66RG
BV327179 S241P6195
AF148494 Drosophila
AR504707 Sequence
AR519989 Sequence
BV267418 S235P6522
AF043416 Schistoso
AF043409 Schistoso
AF043419 Schistoso
AY434012 Schistoso
AY434010 Schistoso
AY434015 Schistoso
AF043417 Schistoso
AF043418 Schistoso
CQ576746 Sequence
CQ575117 Sequence
AC116005 Homo sapi
Z77667 Caenorhabdit
AX251546 Sequence
AC017241 Drosophila
AC101556 Mus muscu
AL353653 Human DNA
BX080806 Dantio rer
AC153681 Bos tauru
AP003841 Oryza sat
AC142019 Rattus no
AC091232 Homo sapi
AP003706 Oryza sat
CP000025 Campyloba
Continuation (151
Continuation (190
Continuation (89 o
Continuation (90 o
Continuation (227
Continuation (275
Continuation (275
Continuation (275
Continuation (34 o
AP006366 Lotus cor
AC090870 Oryza sat
AC105312 Homo sapi
AP003341 Oryza sat
AL078664 Human DNA
AL934662 Mouse DNA
AC165218 Medicago
AC090820 Homo sapi
AP004664 Oryza sat
AP001366 Oryza sat
AP004621 Oryza sat
AP005610 Oryza sat
AC091514 Rattus no
AC091122 Oryza sat
AL631773 Mouse DNA
AP001383 Oryza sat
BX293558 Mouse DNA
AC110043 Mus muscu
AC154408 Mus muscu

C 93	20.2	80.8	164167	9	AC125533	Mus muscu
C 94	20.2	80.8	164190	8	AL655534	Mouse DNA
C 95	20.2	80.8	164810	14	AC091108	AC091108 Homo sapi
C 96	20.2	80.8	164905	14	CR381544	CR381544 Dario rex
C 97	20.2	80.8	165002	14	AP001565	AP001565 Homo sapi
C 98	20.2	80.8	165339	2	AL355814	AL355814 Homo sapi
C 99	20.2	80.8	165766	15	CEY57A10A	AL117195 Caenorhab
C 100	20.2	80.8	166397	15	AP004706	AP004706 Oryza sat
C 101	20.2	80.8	169102	8	AP005192	AP005192 Oryza sat
C 102	20.2	80.8	169780	8	AL359997	AL359997 Human DNA
C 103	20.2	80.8	171787	8	AL353784	AL353784 Human DNA
C 104	20.2	80.8	173146	8	AC100777	AC100777 Homo sapi
C 105	20.2	80.8	173333	14	AC112522	AC112522 Mus muscu
C 106	20.2	80.8	176104	8	AC007163	AC116002 Homo sapi
C 107	20.2	80.8	176591	14	AC001913	AC007163 Homo sapi
C 108	20.2	80.8	178501	14	AL392165	AP001913 Homo sapi
C 109	20.2	80.8	181896	14	AC165203	AL392165 Homo sapi
C 110	20.2	80.8	187523	14	AC102882	AC165203 Atelerix
C 111	20.2	80.8	187662	14	AC113616	AC102882 Mus muscu
C 112	20.2	80.8	190000	14	AC008084	AC113616 Homo sapi
C 113	20.2	80.8	191575	8	AL355858	AC008084 Homo sapi
C 114	20.2	80.8	191587	14	AL355858	AL355858 Human DNA
C 115	20.2	80.8	192518	14	AC027745	AC162830 Bos tauru
C 116	20.2	80.8	193298	14	AC022672	AC027745 Homo sapi
C 117	20.2	80.8	194005	2	AL033774	AC022672 Homo sapi
C 118	20.2	80.8	194045	2	AL033774	AC010653 Drosophili
C 119	20.2	80.8	194354	9	AC079832	AL033774 Mouse DNA
C 120	20.2	80.8	196150	9	AC147250	AC079832 Mus muscu
C 121	20.2	80.8	198250	14	AC093681	AC147250 Mus muscu
C 122	20.2	80.8	200104	9	AC152981	AC093681 Homo sapi
C 123	20.2	80.8	201528	14	AC141486	AC152981 Mus muscu
C 124	20.2	80.8	202404	9	AC135963	AC141486 Rattus no
C 125	20.2	80.8	202366	14	AC090927	AC135963 Mus muscu
C 126	20.2	80.8	203424	8	AC106884	AC090927 Homo sapi
C 127	20.2	80.8	203831	9	AC093295	AC106884 Homo sapi
C 128	20.2	80.8	206117	9	AC079845	AC093295 Homo sapi
C 129	20.2	80.8	207835	14	AP001392	AC079845 Mus muscu
C 130	20.2	80.8	208921	14	AP001392	AP001392 Homo sapi
C 131	20.2	80.8	210172	14	AC129034	AC129034 Rattus no
C 132	20.2	80.8	211582	14	AC121646	AC129034 Rattus no
C 133	20.2	80.8	214419	9	AC101374	AC121646 Rattus no
C 134	20.2	80.8	215366	9	AL603836	AC101374 Mus muscu
C 135	20.2	80.8	215644	8	AC109329	AL603836 Mouse DNA
C 136	20.2	80.8	220541	14	AC156554	AC109329 Homo sapi
C 137	20.2	80.8	222358	14	AC093855	AC156554 Mus muscu
C 138	20.2	80.8	226367	9	AL627425	AC093855 Rattus no
C 139	20.2	80.8	228369	14	AC153665	AL627425 Mouse DNA
C 140	20.2	80.8	228785	14	AC139983	AC153665 Bos tauru
C 141	20.2	80.8	231868	14	AC118124	AC139983 Rattus no
C 142	20.2	80.8	248526	14	AC095612	AC118124 Rattus no
C 143	20.2	80.8	257992	14	AC121389	AC095612 Rattus no
C 144	20.2	80.8	258727	14	AC103272	AC121389 Rattus no
C 145	20.2	80.8	260548	14	AC103274	AC103272 Rattus no
C 146	20.2	80.8	262288	14	AC103274	AC103274 Rattus no
C 147	20.2	80.8	274417	14	AC110653	AC103274 Rattus no
C 148	20.2	80.8	275123	14	AC114855	AC110653 Rattus no
C 149	20.2	80.8	281349	14	CR847855	AC114855 Rattus no
C 150	20.2	80.8	300395	15	AB017111	CR847855 Homo sapi
						AB017111 Oryza sat

ALIGNMENTS

RESULT 1
AC124305/c
LOCUS AC124305
DEFINITION Homo sapiens chromosome 15, clone RP11-173H16, complete sequence.
ACCESSION AC124305
VERSION AC124305.3
KEYWORDS GI:28261592
SOURCE HTG.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 28143)
Britten, B., Linton, L., Nussbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barron, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Bouhagalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,
Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,
Czako, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Gold, S., Goyette, M., Graham, L.,
Grand, P., Hago, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karakas, A., Kells, C., Labocque, K.,
Lamas, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N.,
Matthews, C., McCarthy, M., McGowan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicoll, R., Norbu, C., Norman, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Punthang, P., Pierre, N.,
Pogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S.,
Theodore, J., Topham, K., Travers, M., Travers, N., Triggillo, J.,
Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
Submitted (14-JUN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 28143)

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (14-JUN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 28143)
Britten, B., Nussbaum, C., Lander, E., Abouelell, A., Allen, N.,
Anderson, S., Arrachchi, H., Barron, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Bouhagalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cooke, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J., S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand, P., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karakas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbutt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, T., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicoll, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Punthang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M.,
Vasilev, H., Venkataraman, V., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
Submitted (14-JUN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 28143)
Britten, B., Nussbaum, C., Lander, E., Abouelell, A., Allen, N.,
Anderson, S., Arrachchi, H., Barron, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Bouhagalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cooke, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J., S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand, P., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karakas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbutt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, T., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicoll, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Punthang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,

TITLE
JOURNAL

COMMENT

Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, U., Toplam, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M. Direct Submission

Submitted (06-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 6, 2003 this sequence version replaced gt:27884912.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27334

Center clone name: 173_H_16

Only the first 28.1 kilobases of this clone are being submitted.
The remainder overlaps accession number AC124997 [MITGR project L27341].

FEATURES

SOURCE

Location/Qualifiers
1. 28143
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-173H16"
/clone_lib="RPC1-11 Human Male BAC"
complement(87..203)
/rpt_family="MIR"
570..803
/rpt_family="LIME2"
1227..1719
/rpt_family="LIME2"
1720..2030
/rpt_family="AluSx"
2031..2144
/rpt_family="LIME2"
complement(2150..2547)
/rpt_family="MLT1D"
complement(2558..2694)
/rpt_family="AluJb"
complement(2718..2806)
/rpt_family="MLT1D"
2809..3000
/rpt_family="LIME2"
complement(3344..3644)
/rpt_family="AluY"
complement(3681..3784)
/rpt_family="MER5A"
3855..3983
/rpt_family="MIR"
3995..4118
/rpt_family="LIME2"
complement(4112..4753)
/rpt_family="L1MD2"
complement(4754..5040)
/rpt_family="AluSx"
complement(5041..5289)
/rpt_family="L1MD2"
complement(5290..5454)
/rpt_family="FRAM"
complement(5456..6359)
/rpt_family="L1MD2"
complement(6416..7355)
/rpt_family="L1MD2"
complement(7352..8222)

repeat_region /rpt_family="L1MD2"
8309..8511
/rpt_family="LIME2"
complement(8526..8638)
repeat_region /rpt_family="AluJb"
8643..8890
repeat_region /rpt_family="LIME2"
8891..9182
/rpt_family="AluSx"
9183..9761
/rpt_family="LIME2"
9762..10067
repeat_region /rpt_family="AluSx"
10068..10315
/rpt_family="LIME2"
10318..10374
/rpt_family="MIR"
complement(10626..10933)
repeat_region /rpt_family="AluSx"
complement(12081..12213)
/rpt_family="FLAM_C"
12527..12562
repeat_region /rpt_family="L2"
12771..12874
/rpt_family="A-rich"
complement(14178..14485)
repeat_region /rpt_family="AluSg"
15119..15264
/rpt_family="MIR"
15640..16059
repeat_region /rpt_family="LTR35"
16060..16105
/rpt_family="T(C)n"
16106..16217
repeat_region /rpt_family="LTR35"
16223..17125
/rpt_family="LIME3A"
complement(17158..17461)
repeat_region /rpt_family="AluSx"
complement(17465..17571)
repeat_region /rpt_family="AluSg/x"
complement(17650..17734)
/rpt_family="BLACkACK"
17745..18059
repeat_region /rpt_family="LIME3A"
18065..18154
/rpt_family="MLT1D"

Query Match 100.0%; Score 25; DB 8; Length 28143;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTGATCTTC 25
Db 27523 AAAAAAAAACTAAAGCTGATCTTC 27499

RESULT 2
AC124311 142745 bp DNA linear PRI 08-JAN-2003
LOCUS AC124311 Homo sapiens chromosome 15, clone RP11-479F18, complete sequence.
DEFINITION AC124311
ACCESSION AC124311.7 GI:27545109
VERSION AC124311.7
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 142745)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 15, clone RP11-479F18
JOURNAL Unpublished

REFERENCE
AUTHORS
2 (bases 1 to 142745)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Baetien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzcerald,M., Gage,D., Galgan,J.,
Gardyna,S., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,
Lamasares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Mathews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunhngang,P., Pierre,N.,
Pollars,V., Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Roettli,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
Direct Submission

TITLE
JOURNAL
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
Direct Submission

REFERENCE
AUTHORS
3 (bases 1 to 142745)
Birtten,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzcerald,M., Gage,D., Galgan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunhngang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zemdek,L., Zimmer,A. and Zody,M.
Submitted (16-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 142745)

TITLE
JOURNAL
Submitted (16-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS
5 (bases 1 to 142745)
Birtten,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzcerald,M., Gage,D., Galgan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Mathews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunhngang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL
Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS
5 (bases 1 to 142745)
Birtten,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barna,N., Baetien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzcerald,M., Gage,D., Galgan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Mathews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunhngang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT
JOURNAL
TITLE
JOURNAL
Submitted (08-JUN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 8, 2003 this sequence version replaced gi:24371476.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27342
Center clone name: 479_F_18

FEATURES
source
Only the final 142.75 kb of this clone are being submitted.
The remainder of the clone is overlapped by accession number
AC124997 [WIGR project L27341].
Location/Qualifiers
1. 142745
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-479F18"
/clone_lib="RP11-11 Human Male BAC"
repeat_region
complement(702..1005)
/rpt_family="AluX"
repeat_region
complement(2557..2848)
/rpt_family="AluSc"
complement(2927..3001)
/rpt_family="MIR"
repeat_region
3058..3081
/rpt_family="AT_rich"
repeat_region
3157..3458
/rpt_family="AluDo"
repeat_region
3661..3780
/rpt_family="AluDo"
repeat_region
4078..4210
/rpt_family="MIR3"
repeat_region
4329..4385
/rpt_family="CA)n"
repeat_region
4597..4623
/rpt_family="CA)n"
repeat_region
4637..4746
/rpt_family="U2"
repeat_region
4789..5137
/rpt_family="MLT1A1"
complement(5302..5416)
/rpt_family="MIR"
repeat_region
7580..7600
/rpt_family="GC_rich"
repeat_region
7852..7963
/rpt_family="CCG)n"

```

repeat_region      8418..8439
                    /rpt_family="GC-rich"
repeat_region      9209..9248
                    /rpt_family="(TG)n"
repeat_region      complement(9421..9566)
                    /rpt_family="PLAM C"
repeat_region      complement(10679..10764)
                    /rpt_family="L2"
repeat_region      complement(10852..11162)
                    /rpt_family="ALUy"
repeat_region      11503..11541
                    /rpt_family="MIR"
repeat_region      11545..11567
                    /rpt_family="(TAG)n"
repeat_region      11755..12354
                    /rpt_family="L2"
repeat_region      complement(12348..12449)
                    /rpt_family="L2"
repeat_region      12550..12719
                    /rpt_family="MLT1J"
repeat_region      12720..13106
                    /rpt_family="MLT1A1"
repeat_region      13107..13166
                    /rpt_family="MLT1J"
repeat_region      complement(13271..13328)
                    /rpt_family="L2"
repeat_region      13528..13676
                    /rpt_family="MIR"
repeat_region      13700..13816
                    /rpt_family="MIR"
repeat_region      complement(13893..14396)
                    /rpt_family="L2"
repeat_region      14387..14420
                    /rpt_family="MTR"
repeat_region      14421..14452
                    /rpt_family="L2"
repeat_region      14566..14649
                    /rpt_family="L2"
repeat_region      14930..15139
                    /rpt_family="L2"

Query Match      100.0% Score 25; DB 8; Length 142745;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTGATCTTC 25
DB      99820 AAAAAAAAACTAAGCTGATCTTC 99844

RESULT 3
AC124997      179810 bp DNA linear PRI 27-AUG-2002
DEFINITION    Homo sapiens chromosome 15, clone RP11-466L14, complete sequence.
ACCESSION     AC124997
VERSION       AC124997.4 GI:22507185
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 179810)
Birken,B., Nuebaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-466L14
Unpublished
2 (bases 1 to 179810)
Birken,B., Lincon,L., Nuebaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Charazi,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,

```

```

TITLE
JOURNAL
COMMENT

Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,
Lamazaras,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N.,
Matthews,C., McCarthy,M., McEwan,P., McErmann,K., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,Y., Murphy,T., Naylor,J., Nguyen,C.,
Nicoli,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Teefaye,S.,
Theodore,J., Topham,K., Travers,M., Trivis,N., Triggillo,D.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 179810)
Birken,B., Nuebaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhgalter,B.,
Camarata,J., Chang,J., Charazi,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,Y.,
Murphy,T., Naylor,J., Nguyen,C., Nicoli,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Teefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 179810)
Birken,B., Nuebaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouhgalter,B.,
Camarata,J., Chang,J., Charazi,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,Y.,
Murphy,T., Naylor,J., Nguyen,C., Nicoli,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Teefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 27, 2002 this sequence version replaced gi:22123135.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

```

Center project name: L27341
Center clone name: 466_L_14
-----
FEATURES
Source
Location/Qualifiers
1..179810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-466L14"
/clone_lib="RP11 Human Male BAC"
complement(1063..1359)
/rpc_family="AluSc"
complement(1617..1765)
/rpc_family="AluY"
repeat_region
1803..2129
/rpc_family="AluYb"
repeat_region
3130..3163
/rpc_family="T(G) n"
3463..3540
/rpc_family="L2"
complement(4827..5124)
/rpc_family="AluSg"
complement(5685..5891)
/rpc_family="MER112"
6973..7045
/rpc_family="MIR"
complement(7140..7356)
/rpc_family="MIR"
7375..7458
/rpc_family="L1ME4A"
7609..7735
/rpc_family="AluYb"
7736..7786
/rpc_family="T(AAA)n"
7787..7935
/rpc_family="AluYb"
7963..7991
/rpc_family="AT rich"
complement(8342..8655)
/rpc_family="AluY"
9177..9488
/rpc_family="AluY"
9675..9757
/rpc_family="MIR"
10469..10503
/rpc_family="(CA)n"
10577..10645
/rpc_family="MER5A"
10651..10677
/rpc_family="(TTTA)n"
10874..10965
/rpc_family="(TATATG)n"
10966..11000
/rpc_family="(TTAAA)n"
11001..11050
/rpc_family="(TATATG)n"
11247..11385
/rpc_family="MER104"
11594..11683
/rpc_family="MIR3"
16212..16340
/rpc_family="MLT1U1"
complement(17161..17460)
/rpc_family="AluSc"
20047..20355
/rpc_family="AluYb"
21951..22079
/rpc_family="MIR3"
complement(22137..22374)
/rpc_family="MIR"
repeat_region
22659..22934

```

```

repeat_region
23419..23629
/rpc_family="MIR"
repeat_region
23630..23983
/rpc_family="MER58B"
repeat_region
23984..24101
/rpc_family="THE1A"
repeat_region
24229..24321
/rpc_family="MER58B"
repeat_region
25716..26029
/rpc_family="GA-rich"
complement(24880..24968)
/rpc_family="L1MA4A"
26166..27037
/rpc_family="AluY"
27038..27341
/rpc_family="L1MB1"
27342..27860
/rpc_family="AluSg"
27861..28169
/rpc_family="L1MB1"
28170..29250
/rpc_family="AluSc"
29251..30314
/rpc_family="L1MB1"
30315..30750
/rpc_family="L1MC1"
complement(30765..30829)
/rpc_family="T199ers5"
30830..30923
/rpc_family="L1MC1"
30924..31228
/rpc_family="AluY"

Query Match 100.0%; Score 25; DB 8; Length 179810;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTTC 25
Db 172747 AAAAAAAAACTAAGCTGATCTTC 172771

RESULT 4
AC004738 147404 bp DNA linear PRI 25-NOV-1998
LOCUS Homo sapiens Chromosome 15q11-q13 PAC clone pDJ351h23 from the
ACCESSION Prader-Willi/Angelman Syndrome region, complete sequence.
AC004738
VERSION AC004738.1 GI:3927853
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 147404)
Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
Buetner,J., Buetner,R., Card,P., desailboet,C., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
HTGS Submission
Unpublished
2 (bases 1 to 147404)
Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
Buetner,J., Buetner,R., Card,P., desailboet,C., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
HTGS Submission
Unpublished
JOURNAL REFERENCE
AUTHORS
TITLE
Direct Submission

```



```

repeat_region      110321..110606
                    /rpt_family="Alu"
repeat_region      111104..111155
                    /rpt_family="MER42"
repeat_region      111716..112012
                    /rpt_family="Alu"
repeat_region      112672..112802
                    /rpt_family="Alu"
repeat_region      112873..113204
                    /rpt_family="MT1"
repeat_region      complement(113370..
                    /rpt_family="Alu"
repeat_region      119482..119763
                    /rpt_family="MER2"
repeat_region      120050..120270
                    /rpt_family="Alu"
repeat_region      complement(120965..
                    /rpt_family="Alu"
repeat_region      complement(121456..
                    /rpt_family="MT1"
repeat_region      124613..124768
                    /rpt_family="LTR9"
repeat_region      complement(125001..
                    /rpt_family="LTR12"
repeat_region      125759..126009
                    /rpt_family="LTR9"

```

Query Match 93.6%; Score 23.4; DB 8; Length 147404;
 Best Local Similarity 96.0%; Pred. No. 53;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTTC 25
 Db 87805 AAAAAAAAACTAAAGCTTGATCTTC 87829

RESULT 5
 BV456584/c 628 bp DNA linear STS 06-APR-2005
 LOCUS grf79804.h1 Clint Pan troglodytes versus STS genomic, sequence
 DEFINITION tagged site.
 ACCESSION BV456584
 VERSION BV456584.1 GI:62216012
 KEYWORDS STS.
 SOURCE Pan troglodytes versus
 ORGANISM Pan troglodytes versus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Pan.
 1 (bases 1 to 628)
 Mikelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
 Jaffe,D.B.
 Initial Sequence of the Chimpanzee Genome and Comparison with the
 Human Genome
 Unpublished (2005)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Contact: Michael C. Zody
 Broad Institute of MIT and Harvard
 320 Charles Street, Cambridge, MA 02141, USA
 Tel: 6172580933
 Fax: 6172580903
 Email: mczody@broad.mit.edu
 Primer A: No sequence submitted
 Primer B: No sequence submitted
 STS size: 628
 Protocol:
 23,021,928 chimpanzee whole genome shotgun reads were aligned to
 the Human genome NCBI
 Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,
 including Clint (Pan
 troglodytes verus), 3 other Pan troglodytes versus chimps
 (Donald, Karlien, Yvonne), 3 Pan
 troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps

FEATURES

source

of unknown origin
 (Gen,Unknown Chimp). Common names: Pan troglodytes versus is the
 western chimp and Pan
 troglodytes troglodytes is the central chimp. To be included in
 chimpanzee SNP discovery, a
 read must be at least 500bp in length, at least 50% of its base
 calls must have Phred
 score >= 20, at least 30% of its base calls must satisfy
 SNOS(30,25) (single strand NOS, the
 base in question has Phred score >= 30, the surrounding 10 bases in
 the read have Phred
 score >= 25), and the read must have at least 200 bp SNOS(30,25)
 bases. Reads not uniquely
 placed in the genome and read pairs whose two ends were not
 consistently placed were
 discarded. After above filtering, NOS(30,25) standard was applied
 to all pairs of
 overlapping reads to call NOS bases and SNPs. Alignments (between
 two reads) with less
 than 100 NOS bases or with SNP rate > 0.01 were discarded. To
 exclude alignment between two
 copies of a single read, comparisons between two reads that share
 95% of their genome
 alignments (>=95% bases of read A and >=95% bases of read B were
 placed at the same locus
 of human genome) were discarded.
 Location/Qualifiers
 1..628

STS

ORIGIN

Query Match 89.6%; Score 22.4; DB 10; Length 628;
 Best Local Similarity 95.8%; Pred. No. 7.4e+02;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 24
 Db 532 AAAAAAAAACTAAAGCTTGATCTT 509

RESULT 6
 BV491883 724 bp DNA linear STS 07-APR-2005
 LOCUS S221P61080RH1.T0 Yvonne Pan troglodytes troglodytes STS genomic,
 DEFINITION sequence tagged site.
 ACCESSION BV491883
 VERSION BV491883.1 GI:62335553
 KEYWORDS STS.
 SOURCE Pan troglodytes troglodytes
 ORGANISM Pan troglodytes troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Pan.
 1 (bases 1 to 724)
 Mikelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
 Jaffe,D.B.
 Initial Sequence of the Chimpanzee Genome and Comparison with the
 Human Genome
 Unpublished (2005)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Contact: Michael C. Zody
 Broad Institute of MIT and Harvard
 320 Charles Street, Cambridge, MA 02141, USA
 Tel: 6172580933
 Fax: 6172580903
 Email: mczody@broad.mit.edu
 Primer A: No sequence submitted
 Primer B: No sequence submitted

STS size: 724

Protocol:

23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlien, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score >= 20, at least 30% of its base calls must satisfy SNQs(30,25) (single strand NQs, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred score >= 25), and the read must have at least 200 bp SNQs(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQs(30,25) standard was applied to all pairs of overlapping reads to call NQs bases and SNPs. Alignments (between two reads) with less than 100 NQs bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>95% bases of read A and >95% bases of read B were placed at the same locus of human genome) were discarded.

FEATURES

source

1..724

/organism="Pan troglodytes troglodytes"

/mol_type="genomic DNA"

/sub_species="troglodytes"

/db_xref="taxon:37011"

/clone_1ib="Yvonne"

<1..>724

STS

ORIGIN

Query Match

Best Local Similarity . 89.6%; Score 22.4; DB 10; Length 724;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy

1 AAAAAAAAACTAAGCTTGATCTT 24

Db

38 AAAAAAAAACTAAGCTTGATCTT 61

RESULT 7

LOCUS

AK221711/C

DEFINITION

Arabidopsis thaliana gene for hypothetical protein, complete cds,
clone: RAF16-77-D23.

ACCESSION

AK221711

VERSION

AK221711.1

KEYWORDS

FLI CDNA.

SOURCE

Arabidopsis thaliana

ORGANISM

Arabidopsis thaliana (thale cress)

REFERENCE

1

AUTHORS

Toroki, Y., Seki, M., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Shin-i, T., Nakagawa, M., Sakamoto, N., Oishi, K., Kohara, Y., Kobayashi, M., Toyoda, A., Sakaki, Y., Sakurai, T., Iida, K., Akiyama, K., Satou, M., Toyoda, T., Konagaya, A., Carninci, P., Kawai, D., Hayashizaki, Y. and Shinozaki, K.

TITLE

JOURNAL

REFERENCE

AUTHORS

Large-scale analysis of RIKEN Arabidopsis full-length (RAPL) cDNAs
Unpublished
2 (bases 1 to 1681)
Toroki, Y., Seki, M., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Shin-i, T., Nakagawa, M., Sakamoto, N., Oishi, K., Kohara, Y., Kobayashi, M., Toyoda, A., Sakaki, Y., Sakurai, T., Iida, K., Akiyama, K., Satou, M., Toyoda, T., Konagaya, A., Carninci, P., Kawai, D., Hayashizaki, Y. and Shinozaki, K.
Direct Submission
Submitted (22-MAR-2005) Motoaki Seki, RIKEN Genomic Sciences Center: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: msekis@sc.riken.jp, URL: http://range.gsc.riken.jp/, Tel: 81-45-503-9625, Fax: 81-45-503-9586)

COMMENT

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). This clone is in a modified pBluescript vector. Please visit our web site (http://range.gsc.riken.jp/) for further details.

FEATURES

source

1..1681

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAF16-77-D23"

/ecotype="Columbia"

/note="common name: thale cress"

1..1681

/gene="At3g21820"

61..1482

/gene="At3g21820"

/codon_start=1

/product="hypothetical protein"

/protein_id="BAD95436.1"

/db_xref="GI:62320769"

/translation="MDSVYKTDENFAADVALLAPLPFQIQEYFNKLITSRNCIGIEVNNGTIGKGVYANSEPFDEDELILQDEILVGIQHSNKKVDCVCSFPAFISIEKQI GRKLYFNKLVGSCDDSDSEBECEVKNNGECCGSSSHNTLPBGVSSIMNGEM ALPHTDKFPPLPSPLSCPGGCOAPFCSSCAADNDESSHLCTGERSRSISREBALGE PIKXANDPNDIFLLAAKAIATPILIAIRKLEAHNDKKAKQSPKQSLLEAKPVLG YKRWMDCTALPDVDPTDEGAFRQIKRLACTSLBLKIALFDCEBLAFSLIETIGN IIGMFELNNLIDLIVASPVEDYPLVYIDLPDAKEETETITPFLALGDEYDCCGT AEPFLQSCWNSCCENAKAFKREBDRDQAVIIALRISKNEBVTISYIDELPYKER QALADYGRSCSKCKLEDSSSI"

ORIGIN

Query Match

Best Local Similarity . 89.6%; Score 22.4; DB 15; Length 1681;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy

2 AAAAAAAAACTAAGCTTGATCTTC 25

Db

1646 AAAAAAAAACTAAGCTTGATCTTC 1623

RESULT 8

LOCUS

AY059899

DEFINITION

Arabidopsis thaliana unknown protein (At3g21810; MSD21.12) mRNA,
complete cds.

ACCESSION

AY059899

VERSION

AY059899.1

KEYWORDS

FLI CDNA.

SOURCE

Arabidopsis thaliana

ORGANISM

Arabidopsis thaliana (thale cress)

REFERENCE

1

AUTHORS

Nguyen, M., Karlin-Neumann, G., Southwick, A., Lam, B., Miranda, M., Palm, C. J., Bowser, L., Jones, T., Bann, J., Carninci, P., Chen, H.,

Chen, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A., Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shim, P., Yamada, K., Hayashizaki, K., Ecker, J., Theologis, A. and Davis, R.W.
Direct Submission
Submitted (22-OCT-2001) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
e mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Nguyen, M., Southwick, A., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J., Bower, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shim, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Nguyen, M. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

location/Qualifiers

1. 1839

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="III"

/clone="RAPL08-19-N06"

/note="This clone is in pBluescript"

1. 1839

/gene="At3g21810; MSD21.12"

271. 1584

/gene="At3g21810; MSD21.12"

/codon_start=1

/product="unknown protein"

/protein_id="PAL24381.1"

/db_xref="GI:16649059"

/translation="MSGSSMYKTCILFNKGDGSRNCTFAHGNARRRPGSSFTGRHMSDILDRHNMDSILDRGRGSPSPRDLRSGRVGRSGHNSMPFENRRDXYRNRPRFEDRDYAGCIKYVNRLEDRAEDCRKFRHNNVLEQLKOVENDVRLTDDKLELSEVERKAHEVDLTSRIQLELTQLDRKEDCRITSSKRVKYNRRLAODDLKRSARLQKLNQSLTYLAGSGNNRDVGLIVDEBTNGNNLTACDPHNELQNTSLSRKHVYDQYTTKEPVEDGILGRGEERKVENKRRPCNMILSKSYSEESGAMNDEDTINRSSKEDNKRFRFRIGTSATDQVILSTMAAREPDDVAESEENPEAANGSPILSLPPPPPRDAHVQRNEDVDVNGDVMQKAYDDDSV"

ORIGIN

Query Match 89.6%; Score 22.4; DB 15; Length 1839;
Best Local Similarity 95.8%; Pred. No. 5.2e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTATCTTC 25
|||||
1689 AAAAAAAGAAAGCTTATCTTC 1712
|||||

RESULT 9
AB025634
LOCUS AB025634 59793 bp DNA linear PLN 14-FEB-2004
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MSD21.
ACCESSION AB025634 BA000014
VERSION AB025634.1 GI:4589440

SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCES
AUTHORS
TITLE
JOURNAL
COMMENT

1
Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)

2 (bases 1 to 59793)
Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.
Direct Submission
Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 153-3, Yata, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
Tel:81-438-52-3935, Fax:81-438-52-3934)
Addresses for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see <http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=MSD21>
Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, <http://combio.ornl.gov/Graal-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCP-081.mit.edu/GENSCAN.html>), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremli.ni.zozi.laestate.edu/cgi-bin/sp.cgi>).
Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is M123 and the 3' clone is MEK6.

location/Qualifiers

1. 59793

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/db_xref="taxon:3702"

/chromosome="3"

/clone="MSD21"

/clone_lib="M123" pi"

/ecotype="Columbia"

join(1463..1831,1938..2061,2142..2278,2717..2800,2908..3072)

/note="unlabeled protein product; contains similarity to SGPI monomeric G-protein; GTPase gene id:MSD21.1"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB02832.1"

/db_xref="GI:11994637"

/translation="MAOSCLKIVRINLRNVRRIILRRFTLLMSRIYACTPGKSRRVYLRALPSPVGRSPSPPIPAVVYVGGGGGGEFVRSSVVDNDNSHRSDSDIVSLKISLIGDPEIGTSPRAKYVGEKEVEMLEKGINCTKTYLMGARISYSIWEISGARSRDQIVACKDSYAILFMFDLTSRCTLNVISYQARSSNQAIPTVMVGTKPDEFIQLPIDQLQMTIASQARTYAKALMALTFSSASVYNVNVKIFKFTYAKLDLPWTVERNLTIGEPIDP"

join(5262..5639,5871..5942,6230..6415)

/note="unnamed protein product; gene_id:MSD21.2 unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB02833.1"

/db_xref="GI:11994638"

/translation="MMDTFSGNSYEONHPPHDDIDIAHDHSHGSDHOSRGWTTYLDPESNQYRTTHHEENDHQDSYSYLSLTSYVDAATHAFSGSPFVNPFAKIFGRGRIKICEDSLSDLTASSPVNSPVSHFEHQTPEKTEDIVSSFPWNGIRGMDHQIQIQEGGEQKQVTLRNRLIDGNNNNNNNMIDRSGLCVPIPLSMANNGRC"

complement(join(9671..9768,10108..10700,10977..11579,12203..12609,12726..12755))

/note="gene_id:MSD21.3"

/codon_start=1

CDS

```
/evidence=not_experimental
/product="isocitrate lyase"
/protein_id="BAB02834.1"
/db_xref="GI:11994639"
/translation="MAASFSVPSPIMEEGRPEAEVAEYQVWSSSEPFKLTTRPYTAR
DVALRGLKQGVASNEAMKLMRLTKSHQANGTASRTFGALDPVOVTMMAHGLDTIT
VSGMCCSTHTSTNEGRPLADYPYDVPDKNKEHLPEAQHYDRQREARMSNEER
TKTPFVDYIKPIIADGDTGFGCTTATVYLCFLPERGAAGVHIEDQSSVTKCGMAG
KLVAVSEHINLVARLQFDVWGETVYLVATDVAATLIGSNIADADHOFLIGATN
PSLRKSLSSLSLAEBGTWNGKPAQSLIEDQWLSAGLMTPESEAVVOAIKRMNLNENE
KNORSEWLTTHARVENCLSNBQFALAAAGTDLFEMDMLEPTEGEFRRFGSVAAL
VYRGWAFQIADIIMMETASPDINCTOGAEGIKSTPEVMAVNLSSFPMDASGMT
DOOWEPIRLIARLGCWQFIPLAGFADALVYTPADYARGLMAYVERIQREERT
HGVDTLAHQKSGANYDRYKLTVOGGISSTAMKGGVTEBDFKESWTRPGADMGEG
TSLVVAKSRM"
.complement(join(16643..16838,16962..17206))
/notes="gene_id:MSD21.4"
/codon_start=1
/evidence=not_experimental
/product="dihydroneopterin aldolase-like protein"
/protein_id="BAB02835.1"
/db_xref="GI:11994640"
/translation="MEVYFQRLKLYIVGTESLSPEDKLIIRGLKFYGFHGLPE
EREGLPLVDINIMLSLKKAIESDNLADTVSPADPRLVKVIYVGGPPRNLYETVAD
IASEMLETPKNIVIRVKGKPNPSLVNSTVDPLGAELEPRKSNH"
.complement(join(19424..20171,20253..20533,20672..20934))
/notes="unnamed protein product; gb|AAJ38255.1
gene_id:MSD21.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02836.1"
/db_xref="GI:11994641"
/translation="MAONKSHLMGFRFTASPORVHLVERILFVAVARKTHAFNSPR
NAIEGLKLGNSQLILTLSCFCFYDHEPESAGFVSLIKHLIRLISAMSMRHLMR
KLTSRQASYSSTSSRNKKLKLDLRKLRPMILKRIENRADYPRKEIYPAVEELLIA
RKNLISNTAALIKVPLVLTCKFCEVFYKESGHLLETCKSYIRGNRNLHVEVPSIN
DILVESYVHLNISOVIRHQRERFDYVRVPIELCCQAGLHPEELIOLSEIHDPN
QISEDDIRSLPAGDLKYVGNALNAMEKRAVKKLLVLYPSKVCRCKEVAVGSGH
KARLCGVFESEMGSTHWEKAGVNDLVEKKWHRRPDPVVLVDEGRSYGHAAPAI
VSLCHTGAIVYKAVKCKMKPQGLSFSFTNPVNLNET"
21447..22868
/notes="gene_id:MSD21.6"
/codon_start=1
/evidence=not_experimental
/product="UTP-glucose glucosyltransferase"
/protein_id="BAB02837.1"
/db_xref="GI:11994642"
/translation="MKVELVFIPEPGVGHIRATTLAKLIVASDNRLSVTLIIPSRV
SDASSSVYTNSEDRLRYTLIPARQDTDLVSYIDSKRQVAVSVKAGDVSTSDS
RLAGIVMFCSTSMIDIADEFNLISAVIFTSNASYLGQFPHQSVYDEKDLSEFKD
TEMKPDVPTLRQPPAPKCLPSYMLNKKMPYVLGASRPARGKGLVNSRNLAMEEOAL
SPEFGNGNTNIPVYAVGPIMDLSSGDEERKELIHLIKOAPKSVVPLCRSGMCG
PSEBQAREIYALERSGHRFLMSLRASGVKSNPPRPEFTNLSEIILPKGLDTTVE
IGKITSMAPOVDVLSPAIGAVTHCGMNSILLESIMFGVPMAMPITYAQOQNAFHMV
DELGLAAEYKKEKRRDPLVEBEPIYTADEIERGICAMEODSKMKRVMEMDKLHVA
LVDSGSNCALEKKFYQDVVDNVP"
23981..25438
/notes="gene_id:MSD21.7"
/codon_start=1
/evidence=not_experimental
/product="UTP-glucose glucosyltransferase"
/protein_id="BAB02838.1"
/db_xref="GI:11994643"
/translation="MKLELVFIPSPGDGHLRPLVAVAKLHVDRDHLSTITIIIPOMH
GSSSSSSSYTSLSDSEERLSYVLSVTPKPSDSDPTDKHPRVYIDNFKPVKATVE
KITDPGPPDSRRLGFTVDMFCMMHIDVANEFGVPSITFTISNATFJELQVHAYELY
LVNTFAELEPQAMKFFSGVDSPLPVTVYPMNLKINGPSSDDKOSIILAMEEOAP
RKSIVLFCGSGREGQAKELIALERSGHRFWMSLRAPKSGISGPEEFTNLDEP
ILPEGLFERTAEIGKIVGAPQSAIILANDPAGISGFGMNSLTLESLMGVMAWPL
YAEQOVNAFEMVEBELGLAVEVNSFRGPMADDLMTAEIERGICRLMEDSDVRS
RYKENSEKSHVAMDGSSSHVALLKFIQDVTKNIS"
```

```
join(30227..30448,30615..30803,30886..31048,31128..31543)
/notes="gene_id:MSD21.8"
/codon_start=1
/evidence=not_experimental
/product="peroxidase"
/protein_id="BAB02839.1"
/db_xref="GI:11994644"
/translation="MKITWTQNLAVVVVVTVLIGMLRSSEAOLOMNFYAKSCPNAEKI
ISDHIQNIHNGPSLAAPLIRMHFHDCCVRCQDSVLINSTSGANERDAPPLTLRGF
GFERIKALILEPGEKTVYSCADIIALTARDVAVATGSPSWPTGRDGRISNKTAEAT
NNIPPTNSFTTLQGLFNGOGLNDVLVLGSAHIIGSHCSNMNTRLYNFTYTKOD
PSLSQVYANLKKAKCKSLINDNSTLMDPSSPSRFDLSYRVLVKKRGLRPSDALT
TNSATLKIYINDLVNGSEKFKRPAKRAKSNKRGKRVYKTSAGVIRTRCSVAGS"
.complement(31933..33372)
/notes="gene_id:MSD21.9"
/codon_start=1
/evidence=not_experimental
/product="UTP-glucose glucosyltransferase"
/protein_id="BAB02840.1"
/db_xref="GI:11994645"
/translation="MKTELVPFIPSPALSHMATVEMAEQVLDKNDLSITVYIISFSS
KNTSMITSLTNNRLRYEIIISGDOQPTLEKATDSHISLKPVLVDAVAKLVDTSLPD
APRLAGFVDMVCTSMIDVANEFGVPSYLPYTSNAGPILGLLHIOFMVDAEDIYMSSE
LEDDDELVPSLTSRYPYKCLPYIRKSEKMLTFPVQARRRERKGLVNTVPPLEP
QALTFPLSNGNIPRAVPQPLHLKLVNNDYVDKQSEILRWLDEOPSPRSVPLFCGSM
GFGSEQVRETLALDRSGHRLFLMSLRASPNIIAEPEGEFTNLSEIILPEGFRTAN
RGKVGMAEQVAILAKPAIGFVSHCGMNSTLESIMFGVPMAMPITYAQOQNAFEMV
BELGLAVEIKKHMRGDLGLGRSEIYTAESIIEKGIICLMEQSDVAKRVNEISEKCHVA
LMDGSSSETALKRPIQDVTENIAMSSETS"
.complement(join(33809..35357,35613..35761))
/notes="gene_id:MSD21.10"
/codon_start=1
/evidence=not_experimental
/product="UTP-glucose glucosyltransferase"
/protein_id="BAB02841.1"
/db_xref="GI:11994646"
/translation="MGGRFGQAEIRIAKALDRSSHRFLMSLRASSNIPFKESQESMN
IEVLVPSQKQETACCLPEIRRRTRNMKELVLFYPPQIGHLRSTVENAKLVNDETR
LISVITILPFTISEGVGASDYIALSSANNRLREVLSAVDQPIEMTTEIHNKQO
EPKRSYVAKLLEEDSKSDPKLAGPLDMFCSTMSVAVANFGFSPYMFYSAGIL
SAYTHVOMLGVENKRYDSENDYADSEALNPPISRPVPCVCLPALANMMLPVFVN
QARKFRMKGIYNTVATLEPVLKTSLSPTPPYVPGILLHLENQDSDDERBLE
IIRWLDQOPSPSVPLFCGSGMGREBOVREIATLERSGHRFLMSLRASPNITREI
PGEFTNLSEIILPEGFRTKDIQKVIAGAPVAVLANPAGFVTHCGMNSTLESIMF
```

Query Match
Beat Local Similarity 89.6%; Score 22.4; DB 15; Length 59793;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAAGCTTGATCTTC 25
Db 45347 AAAAAAGACTAAAGCTTATCTTC 45370

RESULT 10
AC104103 137278 bp DNA linear ROD 27-NOV-2003
LOCUS Mus musculus BAC clone RP24-37816 from chromosome 17, complete
DEFINITION
ACCESSION AC104103
VERSION AC104103.5 GI:33457243
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 137278)
AUTHORS Nguyen,C. and Bielicki,L.
TITLE The sequence of Mus musculus BAC clone RP24-37816
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 137278)
AUTHORS Wilson,R.

TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 137278)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-2001) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 137278)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 137278)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 137278)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 6, 2003 this sequence version replaced gi:30725975.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Summary Statistics
Center project name: M_BB0378106

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:
The Rpci-24 BAC Library has been constructed by Pieter de Jong and coworkers (http://www.chori.org) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC122502.

FEATURES
source
 1..137278
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="17"
 /map="17"
 /clone="RP24-37816"
 /clone_lib="RPCI-24"
 repeat_region
 235..318
 /rpc_family="L1"
 repeat_region
 592..1235
 /rpc_family="L1"
 repeat_region
 1383..1828
 /rpc_family="L1"
 4312..4662
 /rpc_family="MALR"
 4689..4841
 /rpc_family="MIR"
 5308..5597
 /rpc_family="B4"
 6999..7142
 /rpc_family="B4"
 7230..7945
 /rpc_family="L1"
 9588..9705
 /rpc_family="B4"
 10600..10830
 /rpc_family="MALR"
 10944..10998
 /rpc_family="L2"
 11030..11079
 /rpc_family="ERV1"
 11049..11095
 /rpc_family="ERV1"
 11288..11332
 /rpc_family="MALR"
 11333..11654
 /rpc_family="MALR"
 11655..11861
 /rpc_family="MALR"
 11970..12107
 /rpc_family="CR1"
 15634..16165
 /rpc_family="L1"
 16295..16412
 /rpc_family="MALR"
 16768..16855
 /rpc_family="CR1"
 17520..17605
 /rpc_family="L2"
 17628..17909
 /rpc_family="B4"
 18051..19106
 /rpc_family="L1"
 19360..19460
 /rpc_family="ALU"
 19468..19605
 /rpc_family="MER1_type"
 19606..19750
 /rpc_family="ALU"
 19751..19859
 /rpc_family="MER1_type"
 20055..20414
 /rpc_family="L1"
 20637..20898
 /rpc_family="B4"
 21414..21539
 /rpc_family="L1"
 21628..22512
 /rpc_family="L1"
 22699..23276
 /rpc_family="L1"
 23854..23981
 /rpc_family="ERVK"
 24036..24312
 /rpc_family="L1"
 24507..24612
 /rpc_family="L1"
 24617..24974
 /rpc_family="MALR"
 24975..25531
 /rpc_family="L1"
 25576..25752
 /rpc_family="ERVK"
 26150..26448

repeat_region 1383..1828
repeat_region /rpc_family="L1"
repeat_region 4312..4662
repeat_region /rpc_family="MALR"
repeat_region 4689..4841
repeat_region /rpc_family="MIR"
repeat_region 5308..5597
repeat_region /rpc_family="B4"
repeat_region 6999..7142
repeat_region /rpc_family="B4"
repeat_region 7230..7945
repeat_region /rpc_family="L1"
repeat_region 9588..9705
repeat_region /rpc_family="B4"
repeat_region 10600..10830
repeat_region /rpc_family="MALR"
repeat_region 10944..10998
repeat_region /rpc_family="L2"
repeat_region 11030..11079
repeat_region /rpc_family="ERV1"
repeat_region 11049..11095
repeat_region /rpc_family="ERV1"
repeat_region 11288..11332
repeat_region /rpc_family="MALR"
repeat_region 11333..11654
repeat_region /rpc_family="MALR"
repeat_region 11655..11861
repeat_region /rpc_family="MALR"
repeat_region 11970..12107
repeat_region /rpc_family="CR1"
repeat_region 15634..16165
repeat_region /rpc_family="L1"
repeat_region 16295..16412
repeat_region /rpc_family="MALR"
repeat_region 16768..16855
repeat_region /rpc_family="CR1"
repeat_region 17520..17605
repeat_region /rpc_family="L2"
repeat_region 17628..17909
repeat_region /rpc_family="B4"
repeat_region 18051..19106
repeat_region /rpc_family="L1"
repeat_region 19360..19460
repeat_region /rpc_family="ALU"
repeat_region 19468..19605
repeat_region /rpc_family="MER1_type"
repeat_region 19606..19750
repeat_region /rpc_family="ALU"
repeat_region 19751..19859
repeat_region /rpc_family="MER1_type"
repeat_region 20055..20414
repeat_region /rpc_family="L1"
repeat_region 20637..20898
repeat_region /rpc_family="B4"
repeat_region 21414..21539
repeat_region /rpc_family="L1"
repeat_region 21628..22512
repeat_region /rpc_family="L1"
repeat_region 22699..23276
repeat_region /rpc_family="L1"
repeat_region 23854..23981
repeat_region /rpc_family="ERVK"
repeat_region 24036..24312
repeat_region /rpc_family="L1"
repeat_region 24507..24612
repeat_region /rpc_family="L1"
repeat_region 24617..24974
repeat_region /rpc_family="MALR"
repeat_region 24975..25531
repeat_region /rpc_family="L1"
repeat_region 25576..25752
repeat_region /rpc_family="ERVK"
repeat_region 26150..26448

```

repeat_region      /rpt_family="L1"
                    26463..26564
                    /rpt_family="B01_MM"
repeat_region      /rpt_family="MALR"
                    27043..27324
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    29522..29602
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    29650..29737
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    31264..32451
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    33911..34064
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    34065..34137
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    34138..34405
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    34408..34545
                    /note="Unresolved simple sequence repeat."
unrepeat_region    /rpt_family="L1"
                    35069..35157
                    /rpt_family="L1"
repeat_region      /rpt_family="B4"
                    35158..35337
                    /rpt_family="B4"
repeat_region      /rpt_family="L1"
                    35584..37021
                    /rpt_family="L1"
repeat_region      /rpt_family="MALR"
                    38187..38389
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    38615..38847
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    39437..39751
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    40806..40995
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    41217..41349
                    /rpt_family="MALR"
repeat_region      /rpt_family="MALR"
                    41860..41909
                    /rpt_family="MALR"

```

Query Match 87.2%; Score 21.8; DB 9; Length 137278;
 Best Local Similarity 92.0%; Pred. No. 2e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAACTAAGCTGATCTTC 25

Db 84746 AAAAAAAAACTAAGCTGATCTTC 84770

RESULT 11
 AC145988/c 162223 bp DNA linear PRI 18-DEC-2003
 LOCUS AC145988 162223 bp DNA linear PRI 18-DEC-2003
 DEFINITION Pan troglodytes BAC clone RP43-166E1 from 7, complete sequence.
 ACCESSION AC145988
 VERSION AC145988.3 GI:38424259
 KEYWORDS HTG
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominae; Pan.
 REFERENCE 1 (bases 1 to 162223)
 AUTHORS Hodges, J. and Haglund, K.
 TITLE The sequence of Pan troglodytes BAC clone RP43-166E1
 JOURNAL Unpublished (2001)
 REFERENCE 2 (bases 1 to 162223)
 AUTHORS Wilson, R.
 TITLE Sequencing of Pan troglodytes
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 162223)
 AUTHORS Wilson, R.K.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 162223)

AUTHORS Wilson, R.K.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 162223)
 AUTHORS Wilson, R.K.
 TITLE Direct Submission
 JOURNAL Submitted (19-NOV-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 6 (bases 1 to 162223)
 AUTHORS Wilson, R.
 TITLE Direct Submission
 JOURNAL Submitted (18-DEC-2003) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 19, 2003 this sequence version replaced gi:38154103.
 COMMENT
 Center: Washington University Genome Sequencing Center
 Center code: MUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@wustl.wustl.edu
 Center project name: C_P1166E01

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. Wes Warren,
 Department of Genetics, Washington University, St. Louis MO. For
 additional information about the map position of this sequence, see
 http://genome.wustl.edu

SOURCE INFORMATION:
 The RP43 BAC Library has been constructed by Chung-Li Shu. DNA
 was isolated from white blood cells obtained from a male chimpanzee
 (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The
 clone and detailed information can be obtained from ResGen
 (http://www.resgen.com) or Pieter de Jong and co-workers at
 http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
 This sequence is the entire insert of the clone.

FEATURES
 source
 1..162223
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="7"
 /map="7"
 /clone="RP43-166E1"
 /clone_11b="RP43-166E1"
 92236..92779
 /note="Sequence derived from one plasmid subclone."

ORIGIN

Query Match 87.2%; Score 21.8; DB 8; Length 162223;
 Best Local Similarity 92.0%; Pred. No. 1.9e+02;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AAAAAAAAACTAAGCTGATCTTC 25

Db 149242 AAAAAAAAACTAAGCTGATCTTC 149218

```

RESULT 12
LOCUS      AL358954/c      88311 bp      DNA      linear      HTG 13-JUN-2001
DEFINITION Homo sapiens chromosome 6 clone RP3-33805, 2 unordered pieces.
ACCESSION  AL358954
VERSION     AL358954.9 GI:10186568
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homiidae; Homo.

REFERENCE   1
AUTHORS     Sims,S.
TITLE       Direct Submission
JOURNAL     Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            Requesters: clonerequest@sanger.ac.uk
            On Sep 19, 2000 this sequence version replaced gi:10129499.

COMMENT     ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: dj33805

            ----- Summary Statistics
            Sequencing vector: plasmid, L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 88097 bases at least Q40
            Consensus quality: 88158 bases at least Q30
            Consensus quality: 88168 bases at least Q20
            Insert size: 88211; sum-of-coverage
            Insert size: 155031; 40.3% error; agarose-ftp
            Quality coverage: 6.97x in Q20 bases; sum-of-coverage
            coverage: 4.03x in Q20 bases; agarose-ftp
            -----

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 22000: contig of 22000 bp in length
            * 22001 22100: gap of 100 bp
            * 22101 88311: contig of 66211 bp in length.

FEATURES             Location/Qualifiers
     source            1..88311
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-33805"
                     /clone_11b="RPCI-3"
     misc_feature       1..22000
                     /note="assembly_fragment:01291
                     clone_end:SP6
                     vector_side:left"
     misc_feature       22101..88311
                     /note="assembly_fragment:00403
                     clone_end:T7
                     vector_side:right"

ORIGIN
Query Match      85.6%; Score 21.4; DB 14; length 88311;
Best Local Similarity 95.7%; Pred. NO.3.3e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY      1 AAAAAAAAACTAAAGCTGATCT 23
DB      25301 AAAAAAAAACTAAAGCTTGATCT 25279

RESULT 13
AC093811          115335 bp   DNA       linear    PRI 01-MAR-2002
LOCUS      Homo sapiens BAC clone RP11-355H11 from 4, complete sequence.
DEFINITION AC093811 AC024662
ACCESSION  AC093811.3 GI:16903168
VERSION    HTG.
KEYWORDS   Homo sapiens (human)
SOURCE     Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homnidae; Homo.
REFERENCE  1 (bases 1 to 115335)
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
PUBMED    9847074
REFERENCE  2 (bases 1 to 115335)
AUTHORS   Shah,N. and Haakenson,W.
TITLE     The sequence of Homo sapiens BAC clone RP11-355H11
JOURNAL   Unpublished (2001)
REFERENCE  3 (bases 1 to 115335)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  4 (bases 1 to 115335)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (11-NOV-2001) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  5 (bases 1 to 115335)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  6 (bases 1 to 115335)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Nov 11, 2001 this sequence version replaced gl:15624965.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0355H11
Drafting Center: WIBR
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.
```


LOCUS	AC112917	134089 bp	DNA	linear	PRI 24-SEP-2002
DEFINITION	Homo sapiens X BAC RP11-46C18 (Roswell Park Cancer Institute Human BAC Library) complete sequence.				
ACCESSION	AC112917				
VERSION	AC112917.3 GI:19807697				
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 134089)				
AUTHORS	Muzny D.M., Adams,C., Adio-Oduola,B., Alt-omran,F.R., Allen,C., Alsebrook,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,U., Benton,J., Blumage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowles,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P., Bulay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Demy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Den,A.L., Ding,Y., Dinh,H.H., Douthaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudah,S., Kalseson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,U., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Marindale,A., Martinez,E., Massey,E., Mashiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,E., Mitchell,T., Mohabbet,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguni,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Plimus,E., Pu,L.L., Ruites,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolle,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostlari,N., Sleson,I., Sodergren,E., Sonakke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabac,P., Tamerisa,A., Thomas,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umaní,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Washington,S., Williams,G., Williamson,A., Wlezyk,R., Woodan,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D., Weinstein,G. and Gibbs,R.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
AUTHORS	2 (bases 1 to 134089)				
TITLE	Worley,K.C.				
JOURNAL	Direct Submission				
AUTHORS	Submitted (25-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 134089)				
AUTHORS	Worley,K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	4 (bases 1 to 134089)				
AUTHORS	Worley,K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	5 (bases 1 to 134089)				
AUTHORS	Worley,K.C.				

```

TITLE      Direct Submission
JOURNAL    Submitted (02-APR-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
REFERENCE  6 (bases 1 to 134089)
AUTHORS    Morley, K.C.
TITLE      Direct Submission
JOURNAL    Submitted (24-SEP-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On Mar 29, 2002 this sequence version replaced gi:19172732.
           INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
           gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
    STS are identified using EPCR (Genome Res. 7:541-550) searches
    of a local database that includes entries from dbSTS, GDB, and
    local mapping efforts.
    Repeats are identified using RepeatMasker (A. Smit and P. Green,
    unpublished.) for Human and Mouse sequences.
    Genes and Region of sequence similarity are identified by BLAST
    (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
    EST and cDNA sequences. Genes demonstrate, at least two exons
    flanked by consensus splice sites that maintained sequence
    continuity across the splice junctions. Sequences that are not
    identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality\_info/genbank\_annotation.html.

QUALSTAT-REPORT.
Location/Qualifiers
source      1. 134089
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="X"
            /clone="RP11-46C18"
            1. 1991
            /note="overlaps bases 192861..194851 of clone AC079264"
            /function="clone overlap"
            258..778
            /rpt_family="L1PA16"
            /complement(2591..3075)
            /rpt_family="L1MB5"
            /complement(3353..3495)
            /rpt_family="L1MB5"
            4130..4160
            /rpt_family="(TTTG)n"
            /complement(4166..4313)
            /rpt_family="AluIo"
            /complement(4314..4619)
            /rpt_family="AluIx"
            /complement(4620..4756)
            /rpt_family="AluIo"

```



```
* 30629 30728: gap of 100 bp
* 30729 37323: contig of 655 bp in length
* 37324 37423: gap of 100 bp
* 37424 42575: contig of 5152 bp in length
* 42576 42675: gap of 100 bp
* 42676 79708: contig of 37033 bp in length
* 79709 79808: gap of 100 bp
* 79809 86724: contig of 6916 bp in length
* 86725 86824: gap of 100 bp
* 86825 109637: contig of 22812 bp in length
* 109637 109736: gap of 100 bp
* 109737 139954: contig of 30218 bp in length
* 139955 140054: gap of 100 bp
* 140055 145722: contig of 5668 bp in length.
FEATURES
    source
        1..145722
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="RP11-21H2"
            /clone_lib="RPC1-11 Human Male BAC"
            1..10503
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
            10504..10603
                /estimated_length=100
            10604..12771
                /note="assembly_fragment"
            12772..12871
                /estimated_length=100
            12872..14989
                /note="assembly_fragment"
            14990..15089
                /estimated_length=100
            15090..16911
                /note="assembly_fragment"
            16912..17011
                /estimated_length=100
            17012..19684
                /note="assembly_fragment"
            19685..19784
                /estimated_length=100
            19785..22088
                /note="assembly_fragment"
            22089..22188
                /estimated_length=100
            22189..25445
                /note="assembly_fragment"
            25446..25545
                /estimated_length=100
            25546..30628
                /note="assembly_fragment"
            30629..30728
                /estimated_length=100
            30729..37323
                /note="assembly_fragment"
            37324..37423
                /estimated_length=100
            37424..42575
                /note="assembly_fragment"
            42576..42675
                /estimated_length=100
            42676..79708
                /note="assembly_fragment"
            79709..79808
                /estimated_length=100
            79809..86724
                /note="assembly_fragment"
            86725..86824
                /estimated_length=100
            86825..109636
                /note="assembly_fragment"
            /note="assembly_fragment"
```

```
gap 109637..109736
misc_feature /estimated_length=100
misc_feature 109737..139954
gap /note="assembly_fragment"
gap 139955..140054
misc_feature /estimated_length=100
misc_feature 140055..145722
misc_feature /note="assembly_fragment"
misc_feature clone_end:SP6
misc_feature vector_side:right"

ORIGIN
Query Match 85.6%; Score 21.4; DB 14; Length 145722;
Best Local Similarity 95.7%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAAGCTTGATCTT 24
DB 46056 AAAAAAAAAATTAAAGCTTGATCTT 46034

RESULT 16
AC073494 160936 bp DNA linear HTG 08-JAN-2003
LOCUS Homo sapiens chromosome X clone RP11-1F14, WORKING DRAFT SEQUENCE,
DEFINITION 11 unordered pieces.
ACCESSION AC073494
VERSION AC073494.9 GI:20335720
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homindaes; Homo.
1 (bases 1 to 160936)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Alabrooks,S.L., Amarantunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dabonne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eathart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,J.,
Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hughes,M., Holloway,C., Hollins,B.,
Homai,F., Howard,S., Huber,J., Huliy,K., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlovic,E., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lonsheed,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabadi,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,B., Nwokweno,S., Ogun,M., Okunnu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pictens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Schere,S., Scott,G., Shen,H., Shoochitari,N., Sisson,I.,
Sodergren,B., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanai,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,K., Wooden,S., Worley,K.,
```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mu.C., Wu.Y., Wu.Y.F., Zhou.J., Zorrilla.S., Nelson.D.,
Weinstock.G. and Gibbs.R.
Direct Submission
Unpublished
2 (bases 1 to 160936)
Worley,K.C.
Direct Submission
Submitted (19-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 160936)
Worley,K.C.
Direct Submission
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:15789212.

----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBSP
Center clone name: RP11-1P14
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-Primer Bodipy: 44% of reads
Chemistry: Dye-terminator Big Dye: 56% of reads
Assembly program: Phrap: version 0.990139
Consensus quality: 152384 bases at least Q40
Consensus quality: 154433 bases at least Q30
Consensus quality: 155875 bases at least Q20
Estimated insert.size: 158401; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 2079: contig of 2079 bp in length
* 2080 2179: gap of unknown length
* 2180 4253: contig of 2074 bp in length
* 4254 4353: gap of unknown length
* 4354 7577: contig of 3224 bp in length
* 7578 7677: gap of unknown length
* 7678 13371: contig of 5694 bp in length
* 13372 13471: gap of unknown length
* 13472 25082: contig of 11611 bp in length
* 25083 25182: gap of unknown length
* 25183 34508: contig of 9326 bp in length
* 34509 34608: gap of unknown length
* 34609 51601: contig of 1693 bp in length
* 51602 51701: gap of unknown length
* 51702 66858: contig of 15157 bp in length
* 66859 66958: gap of unknown length
* 66959 97199: contig of 30241 bp in length
* 97200 97299: gap of unknown length
* 97300 127960: contig of 30661 bp in length
* 127961 128060: gap of unknown length
* 128061 160936: contig of 32876 bp in length.

Location/Qualifiers
1..160936
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
FEATURES
source

```

```

gap      .1      /clone="RP11-1F14"
                2080..2179
gap      /estimated_length=unknown
                4254..4353
gap      /estimated_length=unknown
                7578..7677
gap      /estimated_length=unknown
                13372..13471
gap      /estimated_length=unknown
                25083..25182
gap      /estimated_length=unknown
                34509..34608
gap      /estimated_length=unknown
                51602..51701
gap      /estimated_length=unknown
                66859..66958
gap      /estimated_length=unknown
                97200..97299
gap      /estimated_length=unknown
                127961..128060
gap      /estimated_length=unknown

ORIGIN
Query Match      85.6%; Score 21.4; DB 14; Length 160936;
Best Local Similarity 95.7%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2      AAAAAAAAACTAAGCTTGATCTT 24
          |||||
Db      156880      AAAAAAAAAATTAAGCTTGATCTT 156902

RESULT 17
AC015648/c 166019 bp DNA linear HTG 28-MAR-2000
LOCUS
DEFINITION
AC015648 Homo sapiens clone RP11-46C18, WORKING DRAFT SEQUENCE, 18 unordered
VERSION
AC015648 GI:7329258
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 166019)
Birten,B., Linton,L., Nuebaum,C. and Lander,E.
Homo sapiens, clone RP11-46C18
Unpublished
2 (bases 1 to 166019)
Birten,B., Linton,L., Nuebaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Bouhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArliano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyns,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karataas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGarrk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teague,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 300 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6730920.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

```

```
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L1234
Center clone name: 46 C.18
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-primer-amerzham; 0% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156140 bases at least Q40
Consensus quality: 160439 bases at least Q30
Consensus quality: 162210 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 164319; sum-of-contigs
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bas.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1564: contig of 1564 bp in length
* 1565 1664: gap of 100 bp
* 1665 3109: contig of 1445 bp in length
* 3110 3209: gap of 100 bp
* 3210 6556: contig of 3347 bp in length
* 6557 9048: contig of 2332 bp in length
* 9049 9148: gap of 100 bp
* 9149 10847: contig of 1699 bp in length
* 10848 10947: gap of 100 bp
* 10948 14273: contig of 3336 bp in length
* 14274 14373: gap of 100 bp
* 14374 18490: contig of 4117 bp in length
* 18491 18590: gap of 100 bp
* 18591 22982: contig of 4332 bp in length
* 22983 27831: contig of 4749 bp in length
* 27832 27931: gap of 100 bp
* 27932 32167: contig of 4236 bp in length
* 32168 32267: gap of 100 bp
* 32268 40215: contig of 7948 bp in length
* 40216 40315: gap of 100 bp
* 40316 47861: contig of 7546 bp in length
* 47862 47961: gap of 100 bp
* 47962 62372: contig of 14411 bp in length
* 62373 62472: gap of 100 bp
* 62473 73564: contig of 11092 bp in length
* 73565 73664: gap of 100 bp
* 73665 88229: contig of 14565 bp in length
* 88230 88329: gap of 100 bp
* 88330 108403: contig of 20074 bp in length
* 108404 108503: gap of 100 bp
* 108504 134596: contig of 26093 bp in length
* 134597 134696: gap of 100 bp
* 134697 166019: contig of 31323 bp in length.
*
Location/Qualifiers
1. 166019
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Rp11-46C18"
/clone_1fb="RPC1-11 Human Male BAC"
1. 1564
/note="assembly_fragment"
gap
1565. 1664
/estimated_length=100
misc_feature
1665. 3109
```

```
gap /note="assembly_fragment"
3110. 3209
/estimated_length=100
misc_feature
3210. 6556
/note="assembly_fragment"
6557. 6656
/estimated_length=100
misc_feature
6657. 9048
/note="assembly_fragment"
9049. 9148
/estimated_length=100
misc_feature
9149. 10847
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
10848. 10947
/estimated_length=100
misc_feature
10948. 14273
/note="assembly_fragment"
14274. 14373
/estimated_length=100
misc_feature
14374. 18490
/note="assembly_fragment"
18491. 18590
/estimated_length=100
misc_feature
18591. 22982
/note="assembly_fragment"
22983. 23082
/estimated_length=100
misc_feature
23083. 27831
/note="assembly_fragment"
27832. 27931
/estimated_length=100
misc_feature
27932. 32167
/note="assembly_fragment"
32168. 32267
/estimated_length=100
misc_feature
32268. 40215
/note="assembly_fragment"
40216. 40315
/estimated_length=100
gap
40316. 47861
/note="assembly_fragment"
47862. 47961
/estimated_length=100
misc_feature
47962. 62372
/note="assembly_fragment"
62373. 62472
/estimated_length=100
misc_feature
62473. 73564
/note="assembly_fragment"
73565. 73664
/estimated_length=100
misc_feature
73665. 88229
/note="assembly_fragment"
88230. 88329
/estimated_length=100
misc_feature
88330. 108403
/note="assembly_fragment"
108404. 108503
/estimated_length=100
gap
108504. 134596
/estimated_length=100
misc_feature
134597. 134696
/estimated_length=100
misc_feature
134697. 166019
/note="assembly_fragment"
ORIGIN
Query Match 85.6%; Score 21.4; DB 14; Length 166019;
Best Local Similarity 95.7%; Pred. No. 2.7e+02;
```

Matches	22;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	2	AAAAAAAACTAAAGCTTGATCTT	24						
Db	77013	AAAAAAAAATTAAGCTTGATCTT	76991						
RESULT 18									
LOCUS	AC019179	AC019179	196721 bp	DNA	linear	PRI 09-MAY-2001			
DEFINITION	Homo sapiens BAC clone RP11-240A16	from 4, complete sequence.							
ACCESSION	AC019179								
VERSION	AC019179.4	GI:11120947							
KEYWORDS	HTG;								
SOURCE	Homo sapiens (human)								
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.								
REFERENCE	1 (bases 1 to 196721)								
AUTHORS	Sulston, J.E. and Waterston, R.								
TITLE	Toward a complete human genome sequence								
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)								
PUBMED	9847074								
REFERENCE	2 (bases 1 to 196721)								
AUTHORS	Hakins, R., Maupin, R., Gregory, S., Coblitz, B. and Fleming, A.								
TITLE	The sequence of Homo sapiens BAC clone RP11-240A16								
JOURNAL	Unpublished								
REFERENCE	3 (bases 1 to 196721)								
AUTHORS	Waterston, R.H.								
TITLE	Direct Submission								
JOURNAL	Submitted (30-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA								
REFERENCE	4 (bases 1 to 196721)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA								
REFERENCE	5 (bases 1 to 196721)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA								
COMMENT	On Nov 8, 2000 this sequence version replaced gi:7630907.								
	----- Genome Center								
	Center: Washington University Genome Sequencing Center								
	Center code: WUGSC								
	Web site: http://genome.wustl.edu/gsc								
	Contact: saplens@watson.wustl.edu								
	----- Summary Statistics								
	Center project name: H_NH0240A16								

	NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.								
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.								
	MAPPING INFORMATION:								
	Mapping information for this clone was provided by Dr. John D. McChesron, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc								

SOURCE INFORMATION:
The RPl1-11 human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tateno, M., Carlsene, J. J., and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RPl1-240A16; actual end is at base position 196721 of RPl1-240A16.

The sequence H.NH0240A16 from base position 157677 to 158503 contains a tandem repeat. The assembly is consistent with digest information about the sequence fidelity cannot be guaranteed.

Location/Qualifiers

	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
repeat_region	/clone="RP11-240A16"
	/clone_1fb="RPCI-11"
	3448..1840
repeat_region	/rpt_family="ERV"
	4103..4596
repeat_region	/rpt_family="ERVK"
	5164..5331
repeat_region	/rpt_family="MIR"
	5616..5734
repeat_region	/rpt_family="MIR"
	5861..5911
repeat_region	/rpt_family="Mariner"
	6319..6749
repeat_region	/rpt_family="MALR"
	6727..6776
repeat_region	/rpt_family="MALR"
	8458..8684
repeat_region	/rpt_family="Alu"
	8685..8843
repeat_region	/rpt_family="Alu"
	10345..10912
repeat_region	/rpt_family="MALR"
	10952..11592
repeat_region	/rpt_family="L2"
	14824..15153
repeat_region	/rpt_family="L1"
	16195..16619
repeat_region	/rpt_family="ERVL"
	16649..16822
repeat_region	/rpt_family="MALR"
	17787..17850
repeat_region	/rpt_family="Alu"
	18581..18773
repeat_region	/rpt_family="MBR1_type"
	19637..19697
repeat_region	/rpt_family="L2"
	19796..19971
repeat_region	/rpt_family="Alu"
	19991..20128
repeat_region	/rpt_family="Alu"
	22389..22560
repeat_region	/rpt_family="MIR"
	23974..24287
repeat_region	/rpt_family="Alu"
	25165..25399
repeat_region	/rpt_family="MIR"
	25434..25719
repeat_region	/rpt_family="Alu"
	26014..26157

LTTATIEFWKDLIDVLRIPIDENKEGERIKTDKRLBDTWVAFAPAMINSLFVYVILF
 LQPNODQHLTKRPFQDVALSTOKERNVAVSQEFLMEPLSPISLFLVFPFVWMLQFV
 AMLSHRSYITLHLLSTELHWFPSRRPDMQSENLIERVIAELQKLNDDLR
 AVETDNVSRKRLHNLKARBDTHSVYNDANFNRKRLTLIOGDPVIVSRLSLGDE
 VRRATIRALKTRDSILAEKRSQLOAGDGTGVYNNISGAVNDMSGRASTASYI
 NGVEPAFSDSDDEPRPRRSYVRENTY"

ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 18525;
 Best Local Similarity 91.7%; Pred. No. 9.1e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTT 24
 Db 11024 AAAAAAAAACTAAGCTGATCTT 11001

RESULT 20 AP004509/c 84322 bp DNA linear PLN 22-UTL-2003
 LOCUS Lotu corniculatus var. japonicus genomic DNA, chromosome 6,
 DEFINITION clone:JTL6K17, TM0037a, complete sequence.

ACCESSION AP004509
 VERSION AP004509.1 GI:117736876
 KEYWORDS HTG.
 SOURCE Lotu corniculatus var. japonicus (Lotu japonicus)
 ORGANISM Lotu corniculatus var. japonicus
 Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
 Lotus.

REFERENCE 1 Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. and Tabata, S.
 AUTHORS Structural Analysis of a Lotu japonicus Genome. I. Sequence
 TITLE Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb
 Regions of the Genome
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 84322)
 AUTHORS Nakamura, Y.
 JOURNAL Direct Substitution
 Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research, 1532-3, Yama,
 Kisarazu, Chiba 282-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
 URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935,
 Fax: 81-438-52-3934)

FEATURES
 source Location/Qualifiers
 1. 84322

/organism="Lotu corniculatus var. japonicus"
 /mol_type="genomic DNA"
 /variety="japonicus"
 /db_xref="taxon:34305"
 /chromosome="6"
 /clone="JTL6K17"
 /clone_1fb="JTL1 library"
 /note="TM0037a, a part of TAC clone: TM0037
 synonym: Lotu japonicus"

ORIGIN

Query Match 83.2%; Score 20.8; DB 15; Length 84322;
 Best Local Similarity 91.7%; Pred. No. 5.5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTGATCTT 24
 Db 24439 AAAAAAAAACTAAGCTGATCTT 24416

RESULT 21
 CR382124_07/c
 WPCOMMENT
 Sequence split into 18 fragments LOCUS CR382124 Accession CR382124
 Fragment Name Begin End
 CR382124_00 1 110000

CR382124_01 100001 210000
 CR382124_02 200001 310000
 CR382124_03 300001 410000
 CR382124_04 400001 510000
 CR382124_05 500001 610000
 CR382124_06 600001 710000
 CR382124_07 700001 810000
 CR382124_08 800001 910000
 CR382124_09 900001 1010000
 CR382124_10 1000001 1110000
 CR382124_11 1100001 1210000
 CR382124_12 1200001 1310000
 CR382124_13 1300001 1410000
 CR382124_14 1400001 1510000
 CR382124_15 1500001 1610000
 CR382124_16 1600001 1710000
 CR382124_17 1700001 1715506
 Continuation (8 of 18) of CR382124 from base 700001 (CR382124 Klyveromyces lactis strai

Query Match 83.2%; Score 20.8; DB 15; Length 110000;
 Best Local Similarity 91.7%; Pred. No. 5e+02;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTGATCTTC 25
 Db 80103 AAAAAAAAACTAAGCTGATCTTC 80080

RESULT 22 AC152800/c 157897 bp DNA linear HTG 01-UTL-2005
 LOCUS Bos taurus clone CH240-1116, WORKING DRAFT SEQUENCE, 19 unordered
 DEFINITION pieces.
 ACCESSION AC152800
 VERSION AC152800.3 GI:68226929
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFLOP.
 SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFLOP.
 ORGANISM Bos taurus (cow)
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 Pecora; Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS Muzny, D., Marie, M., Metker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amun, A., Angiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cessat, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Andrade, A., Dederich, D.,
 Delgado, O., Denison, S., Derram, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregorjais, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
 Hollins, B., Howells, S., Hu, Y., Hui, K., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshewa, L., Louisged, H., Lozada, R. J., Lu, X., Ma, J.,
 Maneshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, K., Martin, R., Martiny, E.,
 Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Natr, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Macalenech, O., Okumu, G., Olampunagoon, A., Pal, S., Parks, K.,
 Pasternack, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
 Puzio, M., Quiroz, J., Rachlin, E., Reeves, K., Reiser, M.A., Reigh, R.,
 Rella, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Roese, R., Ruiz, S.,
 Sanders, W., Saverly, G., Scheer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinle, M., Strong, R., Sutton, A., Svetek, A., Taber, P.,
 Taylor, T., Thomas, N., Thomas, S., Tiney, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, Q.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, M., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.
 Direct Submision
 Unpublished
 2 (bases 1 to 157897)
 Morley, K.C.
 Direct Submision
 Submitted (18-NOV-2004) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 157897)
 Cow Genome Sequencing Consortium.
 Direct Submision
 Submitted (01-JUL-2005) Human Genome Sequencing Center, Department

On Jun 26 2005 this sequence version replaced g1:58037818. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: PAPU
 Center clone name: CH240-1116
 ----- Summary Statistics
 Assembly program: Arriba 1.0;
 Consensus quality: 151257 bases at least Q40
 Consensus quality: 152570 bases at least Q30
 Consensus quality: 153714 bases at least Q20
 Estimated insert size: 156424; sum-of-confids estimation
 Quality coverage: 8x in Q20 bases; sum-of-confids estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hpsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	2525: contig of 2525 bp in length
2526	2575: gap of 50 bp

FEATURES	source
*	2576 14761: contig of 12186 bp in length
*	14762 14811: gap of 50 bp
*	14812 16317: contig of 1506 bp in length
*	16318 16417: gap of unknown length
*	16418 36474: contig of 20057 bp in length
*	36475 36524: gap of 50 bp
*	36525 44209: contig of 7665 bp in length
*	44210 44259: gap of 50 bp
*	44260 50710: contig of 6451 bp in length
*	50711 50957: gap of 247 bp
*	50958 52790: contig of 1833 bp in length
*	52791 53670: gap of 880 bp
*	53671 69731: contig of 16061 bp in length
*	69732 69781: gap of 50 bp
*	69782 75001: contig of 5220 bp in length
*	75002 75051: gap of 50 bp
*	75052 77070: contig of 2019 bp in length
*	77071 77120: gap of 50 bp
*	77121 81207: contig of 4087 bp in length
*	81208 81257: gap of 50 bp
*	81258 84306: contig of 3049 bp in length
*	84307 84667: gap of 361 bp
*	84668 96045: contig of 11378 bp in length
*	96046 96095: gap of 50 bp
*	96096 101774: contig of 5679 bp in length
*	101775 101825: gap of 50 bp
*	101826 134661: contig of 32837 bp in length
*	134662 134711: gap of 50 bp
*	134712 153079: contig of 18368 bp in length
*	153080 153179: gap of unknown length
*	153180 154454: contig of 1275 bp in length
*	154454 154554: gap of unknown length
*	154555 156254: contig of 1700 bp in length
*	156255 156354: gap of unknown length
*	156355 157897: contig of 1543 bp in length.
	Location/Qualifiers
	1..157897

```

gap      /organelles="Bos taurus"
gap      /mol_type="genomic DNA"
gap      /db_xref="taxon:9913"
gap      /clone="CH240-11L6"
gap      2526. .2575
gap      /estimated_length=50
gap      14762. .14811
gap      /estimated_length=50
gap      16318. .16417
gap      /estimated_length=unknown
gap      36475. .36524
gap      /estimated_length=50
gap      44210. .44259
gap      /estimated_length=50
gap      50711. .50557
gap      /estimated_length=247
gap      52791. .53670
gap      /estimated_length=880
gap      69732. .69781
gap      /estimated_length=50
gap      75002. .75051
gap      /estimated_length=50
gap      77071. .77120
gap      /estimated_length=50
gap      81208. .81257
gap      /estimated_length=50
gap      84307. .84667
gap      /estimated_length=361
gap      96046. .96095
gap      /estimated_length=50
gap      101775. .101824
gap      /estimated_length=50
gap      134662. .134711
gap      /estimated_length=50
gap      135080. .153179
gap      /estimated_length=unknown

```

gap 154455.154554
gap /estimated_length=unknown
156255.156354
/estimated_length=unknown

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 157897;
Best Local Similarity 91.7%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAAGCTTGATCTTC 25
|||||
85234 AAAAAAAAAATTAAGCTTGAGCTTC 85211
|||||

RESULT 23
CR936342 187909 bp DNA linear HTG 25-APR-2005
LOCUS CR936342/c
DEFINITION Danio rerio clone CH211-215D19, *** SEQUENCING IN PROGRESS ***, 14
unordered pieces.
ACCESSION CR936342
VERSION CR936342.5 GI:62896454
KEYWORDS HTG; HTGS PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 187909)
McLay, K.
Direct Submission
Submitted (24-APR-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 25, 2005 this sequence version replaced gi:62868421.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zc215D19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 182145 bases at least Q40
Consensus quality: 183196 bases at least Q30
Consensus quality: 184102 bases at least Q20
Insert size: 186609; sum-of-contigs
Insert size: 128222; 45.1% error; agarose-fp
Quality coverage: 8.09x in Q20 bases; sum-of-contigs Quality
coverage: 11.78x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 13614: contig of 13614 bp in length
* 13615 13714: gap of 100 bp
* 13715 13707: contig of 3593 bp in length
* 17308 17407: gap of 100 bp
* 17408 22487: contig of 5080 bp in length
* 22488 22587: gap of 100 bp
* 22588 35614: contig of 13027 bp in length
* 35615 35714: gap of 100 bp
* 35715 51458: contig of 15744 bp in length
* 51459 62770: gap of 100 bp
* 62771 62870: contig of 11212 bp in length
* 62871 gap of 100 bp

* 62871 72450: contig of 9580 bp in length
* 72451 72550: gap of 100 bp
* 72551 88008: contig of 15458 bp in length
* 88009 88108: gap of 100 bp
* 88109 90243: contig of 2135 bp in length
* 90244 90344: gap of 100 bp
* 90344 101483: contig of 11140 bp in length
* 101484 101583: gap of 100 bp
* 101584 104058: contig of 2475 bp in length
* 104059 104158: gap of 100 bp
* 104159 107404: contig of 3246 bp in length
* 107405 107504: gap of 100 bp
* 107505 127836: contig of 20332 bp in length
* 127837 127936: gap of 100 bp
* 127937 187909: contig of 59973 bp in length.
Location/Qualifiers
1. 187909
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-215D19"
/clone_1fb="CHORI-211"
1. 13614
/note="assembly_fragment:01008
clone_end:77
vector_side:left"
13715. 17307
/note="assembly_fragment:00023
fragment_chain:1"
17408. 22487
/note="assembly_fragment:00071
fragment_chain:1"
22588. 35614
/note="assembly_fragment:00236
fragment_chain:1"
35715. 51458
/note="assembly_fragment:00629
fragment_chain:1"
51559. 62770
/note="assembly_fragment:00478
fragment_chain:1"
62871. 72450
/note="assembly_fragment:00353
fragment_chain:2"
72551. 88008
/note="assembly_fragment:00805
fragment_chain:2"
88109. 90243
/note="assembly_fragment:00013"
90344. 101483
/note="assembly_fragment:00122"
101584. 104058
/note="assembly_fragment:00035
fragment_chain:3"
104159. 107404
/note="assembly_fragment:00047
fragment_chain:3"
107505. 127836
/note="assembly_fragment:01242
fragment_chain:3"
127937. 187909
/note="assembly_fragment:01477
clone_end:8P6
vector_side:right"

ORIGIN

Query Match 83.2%; Score 20.8; DB 14; Length 187909;
Best Local Similarity 91.7%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 24
|||||

O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (16-JUL-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 16, 2005 this sequence version replaced gi:62868158.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L26442
Center clone name: 362_1_3

FEATURES

source

Location/Qualifiers
1. 214295
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="5"
/map="5"
/clone="RP23-36213"
/clone_lib="RPCI-23 Female Mouse BAC"
complement(5..116)
/rpt_family="ID_B1"
584..635
/rpt_family="(TG)n"
complement(637..776)
/rpt_family="B1_MM"
862..1007
/rpt_family="YREP_MM"
complement(899..1022)
/rpt_family="ID_B1"
complement(1569..1706)
/rpt_family="B4"
1784..1958
/rpt_family="B3A"
complement(2053..2528)
/rpt_family="MERVL_LTR"
complement(2580..2688)
/rpt_family="Lx9"
complement(2802..2949)
/rpt_family="URRIA"
complement(2951..3057)
/rpt_family="Lx9"
3054..3093
/rpt_family="Lx9"
3094..3268
/rpt_family="B2_MM2"
3269..3593
/rpt_family="Lx9"
complement(3627..3764)
/rpt_family="B1_MM"
3822..3942
/rpt_family="B1_MM"
3943..4013
/rpt_family="(GAAA)n"
complement(4059..4247)
/rpt_family="B2_MM2"
complement(4273..4344)
/rpt_family="LIM4"
complement(4367..4844)
/rpt_family="ORR1A-int"

repeat_region complement(4847..5163)
/rpt_family="ORR1A2"
repeat_region complement(5764..5883)
/rpt_family="B1_MM"
repeat_region 5930..6222
/rpt_family="MTE"
repeat_region 6371..6423
/rpt_family="(GGA)n"
repeat_region 6730..6916
/rpt_family="B2_MM1"
repeat_region complement(7271..7392)
/rpt_family="T199er7"
repeat_region 7931..8060
/rpt_family="MIR"
repeat_region complement(8462..8624)
/rpt_family="MSTC"
repeat_region 9282..9372
/rpt_family="PB1D10"
repeat_region 9385..9468
/rpt_family="(CGTG)n"
repeat_region 9543..9608
/rpt_family="B4A"
repeat_region 9781..9989
/rpt_family="B3A"
repeat_region 10011..10155
/rpt_family="RSINE1"
repeat_region complement(10379..10438)
/rpt_family="L2"
repeat_region 10469..10826
/rpt_family="(TG)n"
repeat_region complement(10874..11189)
/rpt_family="B4A"
repeat_region complement(11726..11793)
/rpt_family="RSINE1"
repeat_region complement(12567..12764)
/rpt_family="B2_MM2"
repeat_region 12799..12941
/rpt_family="B1_MM"
repeat_region 12947..12972
/rpt_family="(CAA)n"
repeat_region 13256..13675
/rpt_family="MTC"
repeat_region 13691..13836
/rpt_family="RSINE1"
repeat_region 14094..14125
/rpt_family="AT-rich"
repeat_region 14497..14957
/rpt_family="MLTID"
repeat_region 15840..16040
/rpt_family="(TG)n"
repeat_region 16054..16228
/rpt_family="(TG)n"

Query Match 83.2%; Score 20.8; DB 9; Length 214295;
Best Local Similarity 91.7%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTTAAGCTTGATCTT 24
Db 124276 AAAAAAAAAAGCAAGCTTGATCTT 124299

RESULT 26
AC156538 217253 bp DNA linear HTG 01-JUL-2005
AC156538/c Boe taurus clone CH240-40C7, *** SEQUENCING IN PROGRESS ***; 16
DEFINITION
AC156538
unordered pieces.
AC156538
VERSION AC156538.2 GI:68265381
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Boe taurus (cow)
ORGANISM Boe taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia: Eutheria: Laurasiatheria: Cetartiodactyla: Ruminantia: Pecora: Bovidae: Bovinae: Bos.

1 (bases 1 to 217253)

Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguliano, D., Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Day-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Dlyva, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagay, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, J., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lounsbuwa, L., Loulsegad, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milsavajevic, A., Miner, G., Mijta, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Ngunjiri, M., Norris, S., Nwaekelemeh, O., Okunolu, G., Olamunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., L., Pucato, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reich, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, D., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usumi, K., Vals, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, U., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 217253)

Worley, K. C.

Direct Submission

Submitted (30-JAN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 217253)

Cow Genome Sequencing Consortium.

Direct Submission

Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jun 28, 2005 this sequence version replaced gi:58332918.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: FCNZ
Center clone name: CH240-40C7

----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 21516 bases at least Q40
Consensus quality: 21363 bases at least Q30
Consensus quality: 214829 bases at least Q20
Estimated insert size: 218022; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	2093: contig of 2093 bp in length
2274	gap of 181 bp
2275	5872: contig of 358 bp in length
5873	5922: gap of 50 bp
5923	8612: contig of 2690 bp in length
8613	8662: gap of 50 bp
8663	10874: contig of 2212 bp in length
10875	10974: gap of unknown length
10975	14902: contig of 3928 bp in length
14903	14952: gap of 50 bp
14953	73709: contig of 58757 bp in length
73710	73759: gap of 50 bp
73760	85214: contig of 11455 bp in length
85215	85264: gap of 50 bp
85265	121723: contig of 36459 bp in length
121724	121773: gap of 50 bp
121774	137416: contig of 15643 bp in length
137417	137466: gap of 50 bp
137467	139399: contig of 1933 bp in length
139400	139449: gap of 50 bp
139450	154541: contig of 15092 bp in length
154542	154591: gap of 50 bp
154592	161139: contig of 6548 bp in length
161140	161189: gap of 50 bp
161190	202540: contig of 4151 bp in length
202541	202590: gap of 50 bp
202591	204300: contig of 1710 bp in length
204301	204350: gap of 50 bp
204351	215980: contig of 11630 bp in length
215981	216080: gap of unknown length
216081	217253: contig of 1173 bp in length.

FEATURES

source

1. 217253
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-40C7"
2094. .2274
/estimated_length=181
5873. 5922
/estimated_length=50

```
gap      8613.8662
         /estimated_length=50
gap      10875.10974
         /estimated_length=unknown
gap      14903.14952
         /estimated_length=50
gap      73710.73759
         /estimated_length=50
gap      85215.85764
         /estimated_length=50
gap      121724.121773
         /estimated_length=50
gap      137417.137466
         /estimated_length=50
gap      139400.139448
         /estimated_length=50
gap      154542.154591
         /estimated_length=50
gap      161140.161189
         /estimated_length=50
gap      202541.202590
         /estimated_length=50
gap      204301.204350
         /estimated_length=50
gap      215981.216080
         /estimated_length=unknown

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 217253;
Best Local Similarity 91.7%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCTT 24
Db      206241 ATAAAAAATTAAGCTTGATCTT 206218

RESULT 27
CR931978      218875 bp      DNA      linear      HTG 01-JUN-2005
DEFINITION    Dantio rerio clone CH211-21306, WORKING DRAFT SEQUENCE, 7 unordered
               pieces.
ACCESSION     CR931978
VERSION       CR931978.5 GI:66863539
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE        Dantio rerio (zebrafish)
ORGANISM      Dantio rerio
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
               Cypriniformes; Cyprinidae; Dantio.
               1 (bases 1 to 218875)
REFERENCE     McLaren,S.
AUTHORS       Direct Submission
TITLE         Submitted (30-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
               Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
               zfish-help@sanger.ac.uk Clone requests: clonerquest@sanger.ac.uk
               On Jun 1, 2005 this sequence version replaced gi:56651924.
COMMENT       ----- Genome Center
               Center: Wellcome Trust Sanger Institute
               Center code: SC
               Web site: http://www.sanger.ac.uk
               Contact: zfish-help@sanger.ac.uk
               ----- Project Information
               Center project name: zc21306
               ----- Summary Statistics
               Assembly program: XGAP4; version 4.5
               Chemistry: Dye-terminator; 100% of reads
               Consensus quality: 216050 bases at least Q40
               Consensus quality: 216651 bases at least Q30
               Consensus quality: 217204 bases at least Q20
               Insert size: 218275; sum-of-contigs
               Insert size: 205436; 4.7% error; agarose-fp
               Quality coverage: 11.34x in Q20 bases; sum-of-contigs Quality
```

```
coverage: 12.57x in Q20 bases; agarose-fp
*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 26809: contig of 26809 bp in length
* 26810 26909: gap of 100 bp
* 68268 68268: contig of 4359 bp in length
* 68369 68369: gap of 100 bp
* 68369 167000: contig of 98632 bp in length
* 167001 167100: gap of 100 bp
* 167101 173500: contig of 4250 bp in length
* 173501 174500: gap of 100 bp
* 174501 179079: contig of 7629 bp in length
* 179080 179179: gap of 100 bp
* 179180 181788: contig of 2609 bp in length
* 181789 181888: gap of 100 bp
* 181889 218875: contig of 36987 bp in length.

FEATURES
source
1..218875
/organism="Dantio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-21306"
/clone_id="CHORI-211"
1..26809
/note="assembly fragment:00079
fragment_chain:1
clone_end:T7
vector_side:left"
26910..68268
/note="assembly fragment:01323
fragment_chain:1"
68369..167000
/note="assembly fragment:02099
fragment_chain:1"
167101..173500
/note="assembly fragment:00022"
173501..179079
/note="assembly fragment:00053"
179180..181788
/note="assembly fragment:00036
fragment_chain:2"
181889..218875
/note="assembly fragment:00624
fragment_chain:2
clone_end:SP6
vector_side:right"

ORIGIN
Query Match      83.2%; Score 20.8; DB 14; Length 216875;
Best Local Similarity 91.7%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAAAAAAAACTAAGCTTGATCTT 24
Db      175112 AAAAAAAAACTAAGCTTGATCTT 175089

RESULT 28
AC126071      227741 bp      DNA      linear      ROD 31-AUG-2003
LOCUS         Rattus norvegicus 10 BAC CH230-209B21 (children's Hospital Oakland
DEFINITION    Research Institute) complete sequence.
ACCESSION     AC126071
VERSION       AC126071.5 GI:34368593
KEYWORDS      HTG.
SOURCE        Rattus norvegicus (Norway rat)
```


ORGANISM	COMMENT
Rattus norvegicus	On Aug 31, 2003 this sequence version replaced gi:25138858.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Sequencing is completed to a minimum standard of double strand
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;	coverage with a minimum of 2 clones and 2 reads with no ambiguities
Schirognathi; Muridae; Murioidea; Muridae; Murinae; Rattus.	or 2 chemistries with a minimum of 2 clones and 3 reads with no
1 (bases 1 to 227741)	ambiguities. If the sequence quality does not meet this standard,
Muzny,D,Marie, Metzker,M,Leese, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alebrooks,S, Amin,A, Angiano,D, Anyaldech,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Balwin,D, Bandanaika,D, Barber,M, Barnstead,M, Benahmed,F, Bialo,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guetta,W, Guevara,W, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M, Hollins,B, Howells,S, Hulyk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolyet,A, Karpaty,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshew,L, Louisedge,H, Lozano,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindarne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawhinney,S, McLeod,M, McNeill,T, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwackelme,O, Okunnu,G, Olarunsgoon,A, Pal,S, Parks,K, Pasernak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Plopper,F, Polidexter,A, Popovic,D, Primus,E, Pu,L, Puazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,P, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savary,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smaj,D, Sneed,A, Sodergren,E, Song,X,-Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Uemami,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,O, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wlezyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R,A.	it will be indicated in the annotation. Location/Qualifiers
Unpublished	
2 (bases 1 to 227741)	1. 227741
Worley,K.C.	/organism="Rattus norvegicus"
Direct Submission	/mol_type="genomic DNA"
Submitted (03-JUN-2002) Human Genome Sequencing Center, Department	/db_xref="taxon:10116"
of Molecular and Human Genetics, Baylor College of Medicine, One	/chromosome="10"
Baylor Plaza, Houston, TX 77030, USA	/clone="CH230-209821"
3 (bases 1 to 227741)	complement(397..498)
Rat Genome Sequencing Consortium.	/rpt_family="PB1D9"
Direct Submission	complement(508..683)
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department	/rpt_family="B3"
of Molecular and Human Genetics, Baylor College of Medicine, One	712..854
Baylor Plaza, Houston, TX 77030, USA	/rpt_family="B1_Mur3"
4 (bases 1 to 227741)	1118..1165
Worley,K.C.	/rpt_family="ATYGn"
Direct Submission	complement(1166..1254)
Submitted (31-AUG-2003) Human Genome Sequencing Center, Department	/rpt_family="ID_Rn"
of Molecular and Human Genetics, Baylor College of Medicine, One	complement(1273..1431)
Baylor Plaza, Houston, TX 77030, USA	/rpt_family="RSINB1"
	complement(1999..2088)
	/rpt_family="ID_Rn"
	3586..3656
	/rpt_family="CA_n"
	6415..6510
	/rpt_family="ID_Rn"
	6950..7065
	/rpt_family="PB1D7"
	7404..7437
	/rpt_family="(TG)n"
	7454..7545
	/rpt_family="PB1D10"
	7555..7650
	/rpt_family="ID_Rn"
	7935..7954
	/rpt_family="(A)n"
	7970..8117
	/rpt_family="RSINB1"
	complement(10486..10670)
	/rpt_family="B2_Rn"
	21355..21399
	/rpt_family="GC-rich"
	21520..21644
	/rpt_family="(CCG)n"
	22636..22819
	/rpt_family="B2_Rn"
	22835..22991
	/rpt_family="B4"
	22857..22994
	/rpt_family="B1_Mur3"
	23022..23207
	/rpt_family="B2_Rn"
	23256..23346
	/rpt_family="PB1D7"
	23406..23502
	/rpt_family="PB1D9"
	23623..23654
	/rpt_family="(A)n"
	complement(123929..24117)
	/rpt_family="B2_Rn"
	24167..24305
	/rpt_family="B3A"
	25092..25295
	/rpt_family="B3"
	25322..25459
	/rpt_family="B1_Mur1"
	25460..25498
	/rpt_family="GA-rich"

On Aug 31, 2003 this sequence version replaced gi:25138858. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

FEATURES

source

1. 227741
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/chromosome="10"
/clone="CH230-209821"
complement(397..498)
/rpt_family="PB1D9"
complement(508..683)
/rpt_family="B3"
712..854
/rpt_family="B1_Mur3"
1118..1165
/rpt_family="ATYGn"
complement(1166..1254)
/rpt_family="ID_Rn"
complement(1273..1431)
/rpt_family="RSINB1"
complement(1999..2088)
/rpt_family="ID_Rn"
3586..3656
/rpt_family="CA_n"
6415..6510
/rpt_family="ID_Rn"
6950..7065
/rpt_family="PB1D7"
7404..7437
/rpt_family="(TG)n"
7454..7545
/rpt_family="PB1D10"
7555..7650
/rpt_family="ID_Rn"
7935..7954
/rpt_family="(A)n"
7970..8117
/rpt_family="RSINB1"
complement(10486..10670)
/rpt_family="B2_Rn"
21355..21399
/rpt_family="GC-rich"
21520..21644
/rpt_family="(CCG)n"
22636..22819
/rpt_family="B2_Rn"
22835..22991
/rpt_family="B4"
22857..22994
/rpt_family="B1_Mur3"
23022..23207
/rpt_family="B2_Rn"
23256..23346
/rpt_family="PB1D7"
23406..23502
/rpt_family="PB1D9"
23623..23654
/rpt_family="(A)n"
complement(123929..24117)
/rpt_family="B2_Rn"
24167..24305
/rpt_family="B3A"
25092..25295
/rpt_family="B3"
25322..25459
/rpt_family="B1_Mur1"
25460..25498
/rpt_family="GA-rich"

```

repeat_region      25707..25798
                    /rpt_family="B3A"
repeat_region      26277..26386
                    /rpt_family="B3"
repeat_region      26417..26594
                    /rpt_family="B2_Rn"
repeat_region      26863..26885
                    /rpt_family="AT_rich"
repeat_region      complement(26933..27079)
                    /rpt_family="B1_Mirr4"
repeat_region      28345..28693
                    /rpt_family="B1TR30"
repeat_region      28750..28821
                    /rpt_family="B1F"
repeat_region      28823..28893
                    /rpt_family="ID_Rn"
repeat_region      28894..28914
                    /rpt_family="(TG)n"
repeat_region      29197..29213
                    /rpt_family="AT_rich"
repeat_region      29214..29405
                    /rpt_family="B2_Rn"
repeat_region      29406..29419
                    /rpt_family="AT_rich"
repeat_region      29420..29524
                    /rpt_family="B1_Rn"
repeat_region      29564..29737
                    /rpt_family="B3A"
repeat_region      29797..30099
                    /rpt_family="(CA)n"
repeat_region      complement(30112..30275)
                    /rpt_family="B2_Rn"
repeat_region      30318..30378
                    /rpt_family="(GGA)n"
repeat_region      30379..30748
                    /rpt_family="RMRB"
repeat_region      30772..30867
                    /rpt_family="ID_Rn"
repeat_region      complement(30868..30889)
                    /rpt_family="RMRB"
repeat_region      30890..30981
                    /rpt_family="ID_Rn"
repeat_region      complement(30982..31177)
                    /rpt_family="RMRB"
repeat_region      31178..31281
                    /rpt_family="ID_Rn"
repeat_region      complement(31282..31556)

Query Match      83.2% Score 20.8; DB 9; Length 227741;
Best Local Similarity 91.7%; Pred.No.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTGATCTT 24
Db 23641 AAAAAAAAAAAAGCTTGATCTT 23664

```

JOURNAL REFERENCE AUTHORS

Unpublished
2 (bases 1 to 230880)
Birtten,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,
Chopel,Y., Collymore,A., Cook,A., Cooke,P., Cornu,B.,
DeArrellano,K., Diaz,J.S., Dodge,S., Doolley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,U., Horton,U., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
Maclean,C., MacDonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Menais,L., Mithova,T., Mlanga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,U., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,D., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Stubbs,M., Tajamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (24-MAR-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 230880)

Birtten,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Bouhgalter,B., Camarata,J., Chang,J.,
Chopel,Y., Collymore,A., Cook,A., Cooke,P., Cornu,B.,
DeArrellano,K., Diaz,J.S., Dodge,S., Doolley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,U., Horton,U., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
Maclean,C., MacDonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Menais,L., Mithova,T., Mlanga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,U., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rise,C., Rogov,P., Roman,D., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Stubbs,M., Tajamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.

Direct Submission
Submitted (22-APR-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
On Apr 22, 2005 this sequence version replaced gi:62360737.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Broad Institute of MIT and Harvard
Center code: W1BR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@broad.mit.edu
Project Information
Center project name: L31328
Center clone name: 77_K_13

NOTE: This is a 'working draft' sequence. It currently
consists of 11 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 63601: contig of 63601 bp in length

```

* 63602 63701: gap of unknown length
* 63702 72918: contig of 9217 bp in length
* 72919 73018: gap of unknown length
* 73019 97625: contig of 24607 bp in length
* 97626 97725: gap of unknown length
* 97726 101462: contig of 3737 bp in length
* 101463 101562: gap of unknown length
* 101563 128930: contig of 27368 bp in length
* 128931 129030: gap of unknown length
* 129031 169922: contig of 40892 bp in length
* 169923 170022: gap of unknown length
* 170023 188971: contig of 18949 bp in length
* 188972 189071: gap of unknown length
* 189072 213374: contig of 24303 bp in length
* 213375 213474: gap of unknown length
* 213475 216601: contig of 3127 bp in length
* 216602 216701: gap of unknown length
* 216702 220589: contig of 3888 bp in length
* 220590 220689: gap of unknown length
* 220690 230880: contig of 10191 bp in length.
Location/Qualifiers
1..230880
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-77K13"
/clone_1fb="RP23-77K13 Female Mouse BAC"
63602..63701
/estimated_length=unknown
gap
72919..73018
/estimated_length=unknown
gap
97626..97725
/estimated_length=unknown
gap
101463..101562
/estimated_length=unknown
gap
128931..129030
/estimated_length=unknown
gap
169923..170022
/estimated_length=unknown
gap
188972..189071
/estimated_length=unknown
gap
213375..213474
/estimated_length=unknown
gap
216602..216701
/estimated_length=unknown
gap
220590..220689
/estimated_length=unknown

```

FEATURES

```

source
1..230880
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-77K13"
/clone_1fb="RP23-77K13 Female Mouse BAC"
63602..63701
/estimated_length=unknown
gap
72919..73018
/estimated_length=unknown
gap
97626..97725
/estimated_length=unknown
gap
101463..101562
/estimated_length=unknown
gap
128931..129030
/estimated_length=unknown
gap
169923..170022
/estimated_length=unknown
gap
188972..189071
/estimated_length=unknown
gap
213375..213474
/estimated_length=unknown
gap
216602..216701
/estimated_length=unknown
gap
220590..220689
/estimated_length=unknown

```

ORIGIN

```

Query Match      83.2%; Score 20.8; DB 14; Length 230880;
Best Local Similarity 91.7%; Pred No. 3.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 AAAAAAAAACTAAAGCTGATCTT 24
Db 61633 AAAAAAAAACTAAAGCTGATCTT 61656

```

```

RESULT 30
ACT61390/c 236253 bp DNA 1linear HTG 01-JUL-2005
LOCUS Bos taurus clone CH240-91L12, *** SEQUENCING IN PROGRESS ***, 30
DEFINITION unnumbered pieces.
ACCESSION ACT61390
VERSION ACT61390.2 GI:68302897
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

```

```

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.

```

REFERENCE

AUTHORS

1 (bases 1 to 236253)

Muzny,D,Marie,, Metker,M, Lee,, Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angilano,D., Anylebech,V., Ayogei,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biewald,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceaar,H., Chen,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Ande,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Baves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,D., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huik,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolyet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundaas,M., Murphy,M., Nait,L., Nankervil,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwackeleleh,O., Okwuon,G., Olarunpaagun,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.D., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svetek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Tjofas,Z., Umanu,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhou,D., von Weinstock,G., and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 236253)

Worley,K.C.
Direct Submission
Submitted (13-MAY-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 236253)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jun 29, 2005 this sequence version replaced gi:6396587.
The sequence in this assembly is a combination of BAC-based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atl/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: FFWB

Center clone name: CH240-91L12

----- Summary Statistics

Assembly program: Aclae 3.0;

Consensus quality: 22495 bases at least Q40

Consensus quality: 22750 bases at least Q30

Estimated insert size: 230049 bases at least Q20

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draift_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 30 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

```

1      4981: contig of 4981 bp in length
*      4982      5031: gap of 50 bp
*      5032      7523: contig of 2492 bp in length
*      7524      7573: gap of 50 bp
*      7574      35183: contig of 27610 bp in length
*      35184      35233: gap of 50 bp
*      35234      50751: contig of 15518 bp in length
*      50752      50801: gap of 50 bp
*      50802      65942: contig of 15141 bp in length
*      65943      65992: gap of 50 bp
*      65993      74714: contig of 8722 bp in length
*      74715      74764: gap of 50 bp
*      74765      77508: contig of 2744 bp in length
*      77509      77558: gap of 50 bp
*      77559      82120: contig of 4562 bp in length
*      82121      118027: contig of 35807 bp in length
*      82221      118027: contig of 35807 bp in length
*      118028      118077: gap of 50 bp
*      118078      130768: contig of 12691 bp in length
*      130769      130818: gap of 50 bp
*      130819      153112: contig of 22294 bp in length
*      153113      153162: gap of 50 bp
*      153163      161471: contig of 8309 bp in length
*      161472      161521: gap of 50 bp
*      161522      174530: contig of 13009 bp in length
*      174531      174580: gap of 50 bp
*      174581      178350: contig of 3770 bp in length
*      178351      178400: gap of 50 bp
*      178401      183715: contig of 5315 bp in length
*      183716      183765: gap of 50 bp
*      183766      185217: contig of 1452 bp in length
*      185218      185267: gap of 50 bp
*      185268      189137: contig of 3870 bp in length
*      189138      189187: gap of 50 bp
*      189188      190997: contig of 1810 bp in length
*      190998      191047: gap of 50 bp
*      191048      206593: contig of 15546 bp in length
*      206594      206693: gap of unknown length
*      206694      211029: contig of 4336 bp in length
*      211030      211079: gap of 50 bp
*      211080      212508: contig of 1429 bp in length
*      212509      212608: gap of unknown length
*      212609      217032: contig of 4424 bp in length

```

```

*      217033      217082: gap of 50 bp
*      217083      220073: contig of 2991 bp in length
*      220074      220398: gap of 325 bp
*      220399      221857: contig of 1459 bp in length
*      221858      222179: gap of 322 bp
*      222180      225156: contig of 2977 bp in length
*      225157      225256: gap of unknown length
*      225257      226324: contig of 1068 bp in length
*      226325      226425: gap of unknown length
*      226426      227750: contig of 1326 bp in length
*      227751      227850: gap of unknown length
*      227851      230239: contig of 2389 bp in length
*      230240      230339: gap of unknown length
*      230340      233635: contig of 3296 bp in length
*      233636      233735: gap of unknown length
*      233736      236253: contig of 2518 bp in length.

```

FEATURES

source

1. .236253

/organism="Bos taurus"

/mol_type="genomic DNA"

/db_xref="taxon:9913"

/clone="CH240-91L12"

4982..5031

/estimated_length=50

7524..7573

/estimated_length=50

35184..35233

/estimated_length=50

50752..50801

/estimated_length=50

65943..65992

/estimated_length=50

74715..74764

/estimated_length=50

77509..77558

/estimated_length=50

82121..82220

/estimated_length=unknown

118028..118077

/estimated_length=50

130769..130818

/estimated_length=50

153113..153162

Query Match 83.2%; Score 20.8; DB 14; Length 236253;
Best Local Similarity 91.7%; Pred. No. 3.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY 1 AAAAAAAAAAAGCTTGATCTT 24

DB 155327 AAAAAAAAAAATAAGCTTGATCTT 155304

RESULT 31

AC103058/c

LOCUS

Rattus norvegicus clone CH230-162K12, WORKING DRAFT SEQUENCE, 2

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 246490)

REFERENCE

AUTHORS

Murphy, D., Marie, Metzger, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooke, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Blewett, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Huily, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A., Karyathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kows, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewe, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, D., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nait, L., Nankervis, O., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelmehe, O., Okunolu, G., Olarnpungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plöpper, F., Polndexter, A., Popovic, D., Primus, B., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.D., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Silter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Vales, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Weinstock, G., and Gibbs, R.A.

TITLE
Unpublished
JOURNAL
2 (bases 1 to 246490)
REFERENCE
Worley, K.C.
AUTHORS
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL
3 (bases 1 to 246490)
REFERENCE
Rat Genome Sequencing Consortium.
AUTHORS
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL

COMMENT
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

FEATURES
source
1..246490
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-162K12"
2528..3189
/note="clone boundary
clone end: Sp6
site: EcoRI
end_sequence: BH270848"
complement(32777..33388)
/note="clone boundary
clone end: T7
site: EcoRI
end_sequence: BH270847"
212226..212325
/estimated_length=unknown
gap
212226..212325
ORIGIN
Query Match 83.2%; Score 20.8; DB 14; Length 246490;
Best Local Similarity 91.7%; Pred. No. 3.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTAAAGCTTGATCTT 24
|||||
Db 207857 AAAAAAAAACTAAAGCTTGATCTT 207834
RESULT 32
AC106703
LOCUS
DEFINITION
Rattus norvegicus clone CH230-213C17, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
ACCESSION
AC106703
VERSION
AC106703.4 GI:25139422
KEYWORDS
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 262198)
Muzny, D., Marie, M., Metker, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFCU
Center clone name: CH230-162K12
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 226413 bases at least Q40
Consensus quality: 228880 bases at least Q30
Consensus quality: 230686 bases at least Q20
Estimated insert size: 237154; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 212225: contig of 212225 bp in length
* 212226 212325: gap of unknown length
* 212326 246490: contig of 34165 bp in length.
Location/Qualifiers

Biwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceaar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Derramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Reyes, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, E., Geer, K., Gill, R., Grady, W., Guerra, W., Guerrero, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, J., Hollins, B., Howells, S., Huix, H., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozano, R.U., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Marqu, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Narkaria, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelam, O., Okwunonu, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, B., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodery, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Slason, I., Sitter, C.D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, T., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Walron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weise, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 262198)
 Worley, K.C.
 Direct Submission
 Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 262198)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 20, 2002 this sequence version replaced gi:23270342.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GOSD
 Center clone name: CH230-213C17
 ----- Summary Statistics -----
 Assembly program: Phrap; version 0.990329
 Consensus quality: 200625 bases at least Q40
 Consensus quality: 205791 bases at least Q30
 Consensus quality: 209308 bases at least Q20
 Estimated insert size: 204934; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES	source
1. 262198	Location/Qualifiers
1. 262198	/organism="Rattus norvegicus"
1. 262198	/mol_type="genomic DNA"
1. 262198	/db_xref="taxon:10116"
1. 262198	/clone="CH230-213C17"
1. 1060	misc_feature
1. 1060	/note="wgs contig"
161087..163793	misc_feature
161087..163793	/note="wgs contig"
229054..230815	misc_feature
229054..230815	/note="wgs contig"
255632..256849	misc_feature
255632..256849	/note="wgs contig"
256850..256949	gap
256850..256949	/estimated_length=unknown
238064..258163	gap
238064..258163	/estimated_length=unknown
259295..259394	gap
259295..259394	/estimated_length=unknown
260775..260874	gap
260775..260874	/estimated_length=unknown
ORIGIN	
Query Match	83.2%; Score 20.8; DB 14; Length 262198;
Best Local Similarity	91.7%; Pred. No. 3,8e+02;
Matches	22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DB	1 AAAAAAAAACTAAGCTTGATCTT 24
224465	AAAAAAAAAAAAAAAAAGCTGATCTT 224488
RESULT 33	
AB017243_1	
WPCOMMENT	
Sequence split into 9 fragments	LOCUS AB017243 Accession AB017243
Fragment Name	Begin End
AB017243_0	1 110000

AE017243_1 100001 210000
AE017243_2 200001 310000
AE017243_3 300001 410000
AE017243_4 400001 510000
AE017243_5 500001 610000
AE017243_6 600001 710000
AE017243_7 700001 810000
AE017243_8 800001 897405
Continuation (2 of 9) of AE017243 from base 100001 (AE017243 Mycoplasma hyopneumoniae J,

Query Match 81.6%; Score 20.4; DB 1; Length 110000;
Best Local Similarity 95.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AAAAAAAAACTTAAGCTTGATC 22
Db 85254 AATATAAACTTAAGCTTGATC 85275

RESULT 34
AE017244_01
WPCOMMENT

Sequence split into 10 fragments LOCUS AE017244 Accession AE017244

Fragment Name Begin End
AE017244_00 1 110000
AE017244_01 100001 210000
AE017244_02 200001 310000
AE017244_03 300001 410000
AE017244_04 400001 510000
AE017244_05 500001 610000
AE017244_06 600001 710000
AE017244_07 700001 810000
AE017244_08 800001 910000
AE017244_09 900001 920079
Continuation (2 of 10) of AE017244 from base 100001 (AE017244 Mycoplasma hyopneumoniae J,

Query Match 81.6%; Score 20.4; DB 1; Length 110000;
Best Local Similarity 95.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTAAGCTTGATC 22
Db 92626 AATATAAACTTAAGCTTGATC 92647

RESULT 35
AE017332_2/c
WPCOMMENT

Sequence split into 9 fragments LOCUS AE017332 Accession AE017332

Fragment Name Begin End
AE017332_0 1 110000
AE017332_1 100001 210000
AE017332_2 200001 310000
AE017332_3 300001 410000
AE017332_4 400001 510000
AE017332_5 500001 610000
AE017332_6 600001 710000
AE017332_7 700001 810000
AE017332_8 800001 892758
Continuation (3 of 9) of AE017332 from base 200001 (AE017332 Mycoplasma hyopneumoniae 23

Query Match 81.6%; Score 20.4; DB 1; Length 110000;
Best Local Similarity 95.5%; Pred. No. 7e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTTAAGCTTGATC 22
Db 56757 AATATAAACTTAAGCTTGATC 56736

RESULT 36
BV354105
LOCUS BV354105 553 bp DNA linear STS 27-JAN-2005

DEFINITION S230P66RG6.T0 Rottweiler Canis familiaris STS genomic, sequence tagged site.
ACCESSION BV354105
VERSION BV354105.1 GI:57607984
KEYWORDS STS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 553)
Lindblad-Toh, K.
TITLE The genome sequence of Canis familiaris
JOURNAL Unpublished (2004)
COMMENT

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903

Email: kers11@genome.wi.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 553
Protocol:

WGS-discovery (WGS):
Paired-end low-coverage whole genome shotgun reads were generated from 9 breeds
(German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador Retriever, English Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese Water Dog -100,000 each)
and five other canids (Chinese, Alaskan, Indian and Spanish Gray Wolf as well as the Californian Coyote).

The WGS reads were placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 863872 reads were annotated as STSs and 485941 SNPs were annotated with alleles from the boxer and the breed or canid from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

WGA-discovery (WGA) of Boxer/Poodle SNPs:
A second set of SNPs was generated using a similar methodology except that the contigs from the 1.5x poodle assembly (Kirkness 2003) were used instead of WGS reads. Since this

sequence lacked base quality scores, arbitrary quality scores of phred 40 were assigned before the poodle sequence was placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated with alleles from the boxer and the poodle. The validation rate for these SNPs was estimated at approximately 78%.

Internal-WGA-discovery (I-WGA):
A third set of SNPs were discovered by comparing reads in the WGA assembly. SNPs were defined as mismatch positions that had a base quality of >= 30 on both reads in a region that aligned without gaps, and with at most one additional mismatch in the ten flanking bases. For each allele, at least one additional read had to confirm it. 731476 SNPs were annotated with alleles between the two boxer alleles. The validation rate for these SNPs was estimated at approximately 78%.

FEATURES

1. 553
Location/Qualifiers
/organism="Canis familiaris"
/mol_type="genomic DNA"

```

/strain="Rottweiler"
/db_xref="taxon:9615"
/map="15 90-468 64409592-64409194"
/clone_lib="Rottweiler"
<1..>553

ORIGIN

Query Match      80.8%; Score 20.2; DB 10; Length 553;
Best Local Similarity 88.0%; Pred. No. 4.7e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 499 AAAAAAAAAAAGCTTGCTTC 523

RESULT 37
BV327179      608 bp      DNA      linear      STS 26-JAN-2005
LOCUS      S241P6195FH6.T0 IndiagrayWolf Canis lupus STS genomic, sequence
DEFINITION      tagged site.
ACCESSION      BV327179
VERSION      BV327179.1 GI:57526350
KEYWORDS      STS.
SOURCE      Canis lupus (gray wolf)
ORGANISM      Canis lupus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 608)
Lindblad-Toh, K.
The genome sequence of Canis familiaris
Unpublished (2004)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersti@genome.wi.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 608
Protocol:
WGS-discovery (WGS):
Paired-end low-coverage whole genome shotgun reads were generated
from 9 breeds
(German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
Retriever, English
Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
Water Dog -100,000 each)
and five other canids (Chinese, Alaskan, Indian and Spanish Gray
Wolf as well as the
Californian Coyote).
The WGS reads were placed uniquely on the Canfam1.0 boxer assembly
and SNP detection was
carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
485941 SNPs were
annotated with alleles from the boxer and the breed or canid from
which the particular
read came. The validation rate for these SNPs was estimated at
approximately 98%.
WGA-discovery (WGA) of Boxer/Poodle SNPs:
A second set of SNPs was generated using a similar methodology
except that the contigs
from the 1.5x poodle assembly (Kirkness 2003) were used instead of
WGS reads. Since this
sequence lacked base quality scores, arbitrary quality scores of
phred 40 were assigned
before the poodle sequence was placed uniquely on the Canfam1.0
boxer assembly and SNP
detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated

```

```

with alleles from the
boxer and the poodle. The validation rate for these SNPs was
estimated at approximately TBD%.
Internal-WGA-discovery (I-WGA):
A third set of SNPs were discovered by comparing reads in the WGA
assembly. SNPs were
defined as mismatch positions that had a base quality of >= 30 on
both reads in a region
that aligned without gaps, and with at most one additional mismatch
in the ten flanking
bases. For each allele, at least one additional read had to confirm
it. 731476 SNPs were
annotated with alleles between the two boxer alleles. The
validation rate for these SNPs
was estimated at approximately TBD%.
Location/Qualifiers
1..608
/organism="Canis lupus"
/mol_type="genomic DNA"
/strain="IndiagrayWolf"
/db_xref="taxon:9612"
/map="15 90-468 64409592-64409194"
/clone_lib="IndiagrayWolf"
<1..>608

ORIGIN

STS
Query Match      80.8%; Score 20.2; DB 10; Length 608;
Best Local Similarity 88.0%; Pred. No. 4.6e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 585 AAAAAAAAAAAGTTGATTTTC 561

RESULT 38
AF148494      617 bp      DNA      linear      INV 10-MAY-2000
LOCUS      Drosophila melanogaster mutator2 protein (mu2) gene, promoter
DEFINITION      sequence and partial cds.
ACCESSION      AF148494
VERSION      AF148494.1 GI:4973422
KEYWORDS
SOURCE
ORGANISM      Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 617)
Kasravi, A., Walter, M.F., Brand, S., Mason, J.M. and Biesemann, H.
Molecular cloning and tissue-specific expression of the mutator2
gene (mu2) in Drosophila melanogaster
Genetics 152 (3), 1025-1035 (1999)
10388821
2 (bases 1 to 617)
Kasravi, A., Walter, M.F., Brand, S., Mason, J.M. and Biesemann, H.
Direct Submission
Submitted (04-MAY-1999) Dev. Biol. Center, Univ. of California,
Irvine, CA 92697, USA
Location/Qualifiers
1..617
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3"
/map="62B11-C1"
1..>617
/gene="mu2"
/feature="mutator2"
1..545
/gene="mu2"
497..503
/gene="mu2"
TATA_signal

```

```

/strain="Rottweiler"
/db_xref="taxon:9615"
/map="15 90-468 64409592-64409194"
/clone_lib="Rottweiler"
<1..>553

ORIGIN

Query Match      80.8%; Score 20.2; DB 10; Length 553;
Best Local Similarity 88.0%; Pred. No. 4.7e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
    |||||
Db 499 AAAAAAAAAAAGCTTGCTTC 523

RESULT 37
BV327179      608 bp      DNA      linear      STS 26-JAN-2005
LOCUS      S241P6195FH6.T0 IndiagrayWolf Canis lupus STS genomic, sequence
DEFINITION      tagged site.
ACCESSION      BV327179
VERSION      BV327179.1 GI:57526350
KEYWORDS      STS.
SOURCE      Canis lupus (gray wolf)
ORGANISM      Canis lupus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 608)
Lindblad-Toh, K.
The genome sequence of Canis familiaris
Unpublished (2004)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersti@genome.wi.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 608
Protocol:
WGS-discovery (WGS):
Paired-end low-coverage whole genome shotgun reads were generated
from 9 breeds
(German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
Retriever, English
Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
Water Dog -100,000 each)
and five other canids (Chinese, Alaskan, Indian and Spanish Gray
Wolf as well as the
Californian Coyote).
The WGS reads were placed uniquely on the Canfam1.0 boxer assembly
and SNP detection was
carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
485941 SNPs were
annotated with alleles from the boxer and the breed or canid from
which the particular
read came. The validation rate for these SNPs was estimated at
approximately 98%.
WGA-discovery (WGA) of Boxer/Poodle SNPs:
A second set of SNPs was generated using a similar methodology
except that the contigs
from the 1.5x poodle assembly (Kirkness 2003) were used instead of
WGS reads. Since this
sequence lacked base quality scores, arbitrary quality scores of
phred 40 were assigned
before the poodle sequence was placed uniquely on the Canfam1.0
boxer assembly and SNP
detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated

```


	misc_feature	546..>617 /gene="mu2" /product="mutator2 protein" /note="complete mRNA sequence has been submitted under GenBank Accession Number AF108206"
CDS		546 /gene="mu2" /note="putative transcription start" 615..>617 /gene="mu2" /codon_start=1 /product="mutator2 protein" /protein_id="AAD35085.1" /db_xref="GI:4973423" /translation="M"
ORIGIN		
Query Match	80.8%; Score 20.2; DB 2;	Length 617;
Best Local Similarity	88.0%; Pred. No. 4.6e+03;	
Matches	22; Conservative 0; Mismatches 3;	Indels 0; Gaps 0;
QY	1 AAAAAAAAACTAAGCTTGATCTTC 25 379 AAACAAAACCTAAAGCTCGATTTC 403	
RESULT 39		
LOCUS	ARS04707	627 bp DNA linear PAT 22-SEP-2004
DEFINITION	Sequence 9667 from patent US 6703491.	
ACCESSION	ARS04707	
VERSION	ARS04707.1 GI:52440182	
KEYWORDS	.	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified. 1 (bases 1 to 627)	
AUTHORS	Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L., Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.	
TITLE	Drosophila sequences	
JOURNAL	Patent: US 6703491-A 9667 09-MAR-2004; Exelixis, Inc.; South San Francisco, CA	
FEATURES	Location/Qualifiers source 1..627 /organism="unknown" /mol_type="genomic DNA"	
ORIGIN		
Query Match	80.8%; Score 20.2; DB 6;	Length 627;
Best Local Similarity	88.0%; Pred. No. 4.5e+03;	
Matches	22; Conservative 0; Mismatches 3;	Indels 0; Gaps 0;
QY	1 AAAAAAAAACTAAGCTTGATCTTC 25 121 AAACAAAACCTAAAGCTCGATTTC 97	
RESULT 40		
LOCUS	ARS19989	627 bp DNA linear PAT 22-SEP-2004
DEFINITION	Sequence 24949 from patent US 6703491.	
ACCESSION	ARS19989	
VERSION	ARS19989.1 GI:52455464	
KEYWORDS	.	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	Unclassified. 1 (bases 1 to 627)	
AUTHORS	Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L., Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.	
TITLE	Drosophila sequences	
JOURNAL	Patent: US 6703491-A 24949 09-MAR-2004; Exelixis, Inc.; South San Francisco, CA	

FEATURES	Location/Qualifiers
Source	1. .627 /organism="unknown" /mol_type="genomic DNA"
Query Match	80.8%; Score 20.2; DB 6; Length 627;
Best Local Similarity	88.0%; Pred. No. 4.5e+03;
Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy	1 AAAAAAAAACTAAAGCTGATCTTC 25 121 AACCAAACTAAAGCTGATATTC 97
RESULT 41	
LOCUS	BV267418 654 bp DNA linear STS 22-JAN-2005
DEFINITION	S235P8522RG11.T0 ItalianGreyhound Canis familiaris STS genomic.
ACCESSION	BV267418
VERSION	BV267418
KEYWORDS	BV267418.1 GI:57329960
SOURCE	STS.
ORGANISM	Canis familiaris (dog)
	Canis familiaris
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
	Canis.
REFERENCE	1 (bases 1 to 654)
AUTHORS	Lindblad-Toh,K.
TITLE	The genome sequence of Canis familiaris
JOURNAL	Unpublished (2004)
COMMENT	Contact: Kerstin Lindblad-Toh Whitehead Institute for Biomedical Research, Center for Genome Research 320 Charles Street, Cambridge, MA 02141, USA Tel: 6172521477 Fax: 6172580903 Email: kersti@genome.wi.mit.edu Primer A: No sequence submitted Primer B: No sequence submitted STS size: 654 Protocol: WGS-discovery (WGS): Paired-end low-coverage whole genome shotgun reads were generated from 9 breeds (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador Retriever, English Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese Water Dog -100,000 each) and five other canids (Chinese, Alaskan, Indian and Spanish Gray Wolf as well as the Californian Coyote). The WGS reads were placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 863872 reads were annotated as SNPs and 455941 SNPs were annotated with alleles from the boxer and the breed or canid from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%. WGA-discovery (WGA) of Boxer/Poodle SNPs: A second set of SNPs was generated using a similar methodology except that the contigs from the 1.5x poodle assembly (Kirkness 2003) were used instead of WGS reads. Since this sequence lacked base quality scores, arbitrary quality scores of phred 40 were assigned before the poodle sequence was placed uniquely on the CanFam1.0 boxer assembly and SNP detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated with alleles from the


```
ASTONGAIDLITSSSAPTTWAKALSNANRLITSSPKVKRKQVPLIRQGITPLST
DICEEPLE"
exon      3. .147
          /gene="HSF"
          /number=4
exon      148. .488
          /gene="HSF"
          /number=5
exon      489. .787
          /gene="HSF"
          /number=6
exon      788. .945
          /gene="HSF"
          /number=7
misc_feature 946. .984
          /gene="HSF"
          /note="retained intron"
          /number=7
exon      985. .1041
          /gene="HSF"
          /number=8
exon      1042. .1391
          /gene="HSF"
          /number=9
exon      1092. .>1694
          /gene="HSF"
          /number=10

ORIGIN
Query Match      80.8%; Score 20.2; DB 2; Length 1694;
Best Local Similarity 88.0%; Pred. No. 3.3e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
        |||||
        117 AACAAAGACTAAGCTGATCTTC 141

RESULT 44
SMHSF1      2146 bp      DNA      linear      INV 21-MAY-1998
LOCUS      Schistosoma mansoni heat shock transcription factor, exons 1-4.
ACCESSION      AF043419
VERSION      AF043419.1 GI:2895594
KEYWORDS
SEGMENT
SOURCE      1 of 3
            Schistosoma mansoni
ORGANISM      Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoida; Schistosomatidae; Schistosoma.
REFERENCE
AUTHORS      Lanter, F., Ziv, E., Ram, D. and Schechter, I.
TITLE      1 (bases 1 to 2146)
            factor are expressed during the life cycle of the parasitic
            helminth Schistosoma mansoni
JOURNAL      Eur. J. Biochem. 253 (2), 390-398 (1998)
PUBMED      9654088
REFERENCE
AUTHORS      Lanter, F., Ziv, E., Ram, D. and Schechter, I.
TITLE      2 (bases 1 to 2146)
            Direct Submission
JOURNAL      Submitted (20-JAN-1998) The Weizmann Institute of Science, Rehovot
            76100, Israel
FEATURES
            Location/Qualifiers
            1. .2146
            /organism="Schistosoma mansoni"
            /mol_type="genomic DNA"
            /strain="Puerto Rican"
            /db_xref="taxon:6183"
            /clone="HSF-G1"
            /dev_stage="adult worm"
            1. .934
            promoter
            TATA_signal
            exon      679. .683
            935. .1073

ASTONGAIDLITSSSAPTTWAKALSNANRLITSSPKVKRKQVPLIRQGITPLST
DICEEPLE"
exon      1107. .1215
          /number=1
exon      1249. .1667
          /number=2
exon      1702. .1846
          /number=3
          /number=4

ORIGIN
Query Match      80.8%; Score 20.2; DB 2; Length 2146;
Best Local Similarity 88.0%; Pred. No. 3e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
        |||||
        1816 AACAAAGACTAAGCTGATCTTC 1840

RESULT 45
AY434012      2154 bp      mRNA      linear      INV 19-OCT-2004
LOCUS      Schistosoma mansoni clone CER52 heat shock transcription factor
DEFINITION      (HSF) mRNA, partial cds, alternatively spliced.
ACCESSION      AY434012
VERSION      AY434012.1 GI:40388409
KEYWORDS
SOURCE      Schistosoma mansoni
            Schistosoma mansoni
ORGANISM      Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoida; Schistosomatidae; Schistosoma.
REFERENCE
AUTHORS      Ram, D., Ziv, E., Lanter, F., Lardans, V. and Schechter, I.
TITLE      Stage-specific alternative splicing of the heat-shock transcription
            factor during the life-cycle of Schistosoma mansoni
JOURNAL      Parasitology 129 (5), 587-596 (2004)
PUBMED      154012
REFERENCE
AUTHORS      Ram, D., Ziv, E., Lanter, F., Lardans, V. and Schechter, I.
TITLE      2 (bases 1 to 2154)
            Direct Submission
JOURNAL      Submitted (23-SEP-2003) Immunology, The Weizmann Institute of
            Science, Rehovot 76100, Israel
FEATURES
            Location/Qualifiers
            1. .2154
            /organism="Schistosoma mansoni"
            /mol_type="mRNA"
            /db_xref="taxon:6183"
            /clone="CER52"
            <1. .>2154
            /gene="HSF"
            /note="similar to HSF in GenBank Accession Number
            AF043418"
            <1. .430
            /gene="HSF"
            /note="alternatively spliced"
            /codon_start=2
            /product="heat shock transcription factor"
            /protein_id="AAR85476.1"
            /db_xref="GI:40388410"
            /translation="ISVSPRPITATDPLRLAEYVRHRCNOETLSQOISVLKSNOL
            LYRESDIREHNDQSOGLQITFTLSAPAGSGASVCIGCTRKALPSITPSSRL
            ONKEIKDLRQDFVISTCTVITYVLSFNNQMDRCVLRI"
            <1. .196
            /gene="HSF"
            /number=3
            197. .341
            /gene="HSF"
            /number=4
            342. .620
            /gene="HSF"
            /note="14a-segment; retained intron"
            /number=4
            621. .925
            /gene="HSF"
            /number=5
```

```

exon      837..845
           /gene="HSF"
           /number=AS
exon      926..1224
           /gene="HSF"
           /number=6
exon      1225..1382
           /gene="HSF"
           /number=7
exon      1383..1439
           /gene="HSF"
           /number=8
exon      1440..1789
           /gene="HSF"
           /number=9
exon      1790..2151
           /gene="HSF"
           /number=10

```

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 2154;

Best Local Similarity 88.0%; Pred. No. 3e+03; Mismatches 3; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
         |||||
Db      311 AACAAAGAACTAAGCTGATCTTC 335

```

RESULT 46

AY434010

LOCUS Schistosoma mansonii clone CER3 heat shock transcription factor

DEFINITION (HSF) mRNA, partial cds, alternatively spliced.

ACCESSION AY434010.1 GI:40388405

VERSION

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansonii

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigoididae; Schistosomatidae; Schistosomatidae; Schistosoma.

1 (bases 1 to 2168)

Ram,D., Ziv,E., Lantner,F., Lardans,V. and Schechter,I.

Stage-specific alternative splicing of the heat-shock transcription

factor during the life-cycle of Schistosoma mansonii

2 (bases 1 to 2168)

Ram,D., Ziv,E., Lantner,F., Lardans,V. and Schechter,I.

Direct Submission

Submitted (23-SEP-2003) Immunology, The Weizmann Institute of

Science, Rehovot 76100, Israel

Location/Qualifiers

1..2168

/organism="Schistosoma mansonii"

/mol_type="mRNA"

/db_xref="taxon:6183"

/clone="CER32"

<1..>2168

/gene="HSF"

AF043418

<1..414

/note="alternatively spliced"

/product="heat shock transcription factor"

/protein_id="AA85474.1"

/db_xref="GI:40388406"

/translation="ORPITADPFLAATVRLHRCNORTLSQQLSVLSEMOYLVEL

SDRHHNKGSQLITLFTPLSAPAKERSAVCIQTGRKALPSITPSGSLQNKEL

KLDKDFVISTIVITVLSFTNNQNDRCIVRI"

<1..180

/gene="HSF"

```

exon      /number=3
           181..325
           /gene="HSF"
           /number=4
misc_feature 326..604
           /gene="HSF"
           /note="14a-segment; retained intron"
           /number=4
exon      605..945
           /gene="HSF"
           /number=5
exon      821..865
           /gene="HSF"
           /number=AS
exon      945..1244
           /gene="HSF"
           /number=6
exon      1245..1402
           /gene="HSF"
           /number=7
exon      1403..1459
           /gene="HSF"
           /number=8
exon      1460..1809
           /gene="HSF"
           /number=9
exon      1810..2152
           /gene="HSF"
           /number=10

```

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 2168;

Best Local Similarity 88.0%; Pred. No. 3e+03; Mismatches 3; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy      1 AAAAAAAAACTAAGCTTGATCTTC 25
         |||||
Db      295 AACAAAGAACTAAGCTGATCTTC 319

```

RESULT 47

AY434015

LOCUS Schistosoma mansonii clone WO18 transcription factor (HSF) mRNA,

DEFINITION partial cds, alternatively spliced.

ACCESSION AY434015.1 GI:40388414

VERSION

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansonii

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigoididae; Schistosomatidae; Schistosomatidae; Schistosoma.

1 (bases 1 to 2254)

Ram,D., Ziv,E., Lantner,F., Lardans,V. and Schechter,I.

Stage-specific alternative splicing of the heat-shock transcription

factor during the life-cycle of Schistosoma mansonii

2 (bases 1 to 2254)

Ram,D., Ziv,E., Lantner,F., Lardans,V. and Schechter,I.

Direct Submission

Submitted (23-SEP-2003) Immunology, The Weizmann Institute of

Science, Rehovot 76100, Israel

Location/Qualifiers

1..2254

/organism="Schistosoma mansonii"

/mol_type="mRNA"

/db_xref="taxon:6183"

/clone="WO18"

<1..>2254

/gene="HSF"

/note="similar to HSF in GenBank Accession Number

AF043418"

<1..114

/gene="HSF"

```

exon      /gene="HSF"
           /note="alternatively spliced"
           /codon_start=1
           /product="transcription factor"
           /protein_id="AA085478.1"
           /db_xref="GI:40388415"
           /translation="GTSFHRDGNRLAKELLPLVFKNNLSSFIRQLNMCW"
           <1..106
           /gene="HSF"
           /number=2
           107..139
           /gene="HSF"
           /note="retained intron"
           /number=2
           140..558
           /gene="HSF"
           /number=3
           559..703
           /gene="HSF"
           /number=4
           704..1045
           /gene="HSF"
           /number=5
           1046..1344
           /gene="HSF"
           /number=6
           1345..1501
           /gene="HSP"
           /number=7
           1502..1558
           /gene="HSF"
           /number=8
           1559..1908
           /gene="HSP"
           /number=9
           1909..2254
           /gene="HSF"
           /number=10

```

[illegible]

```

FEATURES
source
location/Qualifiers
1..2261
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/strain="Puerto Rican"
/db_xref="taxon:6183"
/clone="C-250"
/dev_stage="adult worm"
<1..1897
/function="binds to HSE in promoters of Hsp genes"
/note="Hsf2 protein, alternatively spliced"
/codon_start=2
/product="heat shock transcription factor"
/protein_id="AAC39025.1"
/db_xref="GI:2854021"
/translation="PLTKRLVLDDEETNELIYMDPHGTSFHIIRDGNRLAKELLPLVF
KHNLSPIRILQMLWGEGRKIRVYDPSILKSPDHEDFSGPYIRAKNDILLSKIQRT
SNMSPITGSRNOSFGYKVPYVQANSGLSGVSFQPIRTATDFLRALFTVHLRGN
QETLSQIISVLAKENQLLYRELSDLRHHKQSLIOTLFTPLSAFKEGRSAVCIG
QTRKALPITPGSGANLQELKLDLRKQPIIRNSVQPLQVOTSLDLSHKRPED
IRNLASIGITSSGANTDQISDGSVHLPSNLQLTPVLSGNDCECDVQTSIADN
PVNVQDPRYRLSKVDNFPHTNDEEPVGLNVEEDGVANNISIELMPCSRPT
PYCLDDEPGLSDVIPSKLDTGVTDESIIDLILNDSQTEBERGNQNPSSLSKYKE
PIKQKQIRNPRCIIAPRASPQVSTTQIDLSLMEWYQDQSGPPLNDYQNOPE
NMSGSSGLIVGNBIIIPVPEANLNACDVQIFRASEIPGKKLSDEKLTVPATTEAA
STONGAIDLITSSSAPTTPKARLSANNRLITSSPKVKQRPVLLIRKQITPLSTD
FCEEELR"
<1..310
/note="encodes DNA binding domain"
440..622
/note="encodes leucine zipper 123"
878..955
/note="encodes leucine zipper 4"

```

[illegible]

5'UTR
CDS
/db_xref="taxon:6183"
/clone="C-252; C-257; C-517"
/dev_stage="adult worm"
1..78
29..2005
/function="binds to HSE in promoters of Hep genes"
/note="HSF3 protein; alternatively spliced"
/codon_start=1
/product="heat shock transcription factor"
/protein_id="AAC39026.1"
/db_xref="GI:2854023"

misc_feature
misc_feature
misc_feature
misc_feature
3'UTR
ORIGIN
Query Match
Best Local Similarity 80.8%; Score 20.2; DB 2; Length 2369;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAGCTGATCTTC 25
DB 782 AACAAAGAACTAAGCTGATCTTC 806

RESULT 50
COS76746/c
LOCUS
DEFINITION COS76746 4254 bp DNA linear PAT 02-FEB-2004
ACCESSION
VERSION COS76746.1 GI:41639810
KEYWORDS
SOURCE
ORGANISM
Drosophila sp.
Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1
Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
Detection k1te, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
Patent: WO 0171042-A 4504 27-SEP-2001;
PE Corporation (NY) (US)
Location/Qualifiers
1..4254
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN
Query Match
Best Local Similarity 80.8%; Score 20.2; DB 6; Length 4254;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Search completed: December 14, 2005, 11:12:10
Job time : 872.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 13, 2005, 23:35:38 , Search time 203.2 Seconds
(without alignments)
819.967 Million cell updates/sec

Title: US-10-681-773-10

Perfect score: 25
Sequence: 1 aaaaaaaaaaagctgattccttc 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : N_Geneseq_21.*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.4	89.6	1666	13	ADX31754
2	20.2	80.8	4254	4	ABL04842
3	20.2	80.8	6746	4	ABL03756
4	20.2	80.8	6120	4	AA546788
5	20.2	80.8	62154	12	AD097611
6	20.2	80.8	110000	13	ABD32868_3
7	20.2	80.8	235070	11	ACN45174
8	19.8	79.2	13606	4	AA545458
9	19.8	79.2	13606	4	AA546562
10	19.8	79.2	13606	4	ABL33811
11	19.8	79.2	13606	6	ABK28314
12	19.2	76.8	387	5	ABV07913
13	19.2	76.8	400	5	ABV07451
14	19.2	76.8	424	5	ABV04081
15	19.2	76.8	1834	10	ADB87460
16	19.2	76.8	1834	12	ADJ35269
17	19.2	76.8	1834	12	ADJ96887
18	19.2	76.8	1999	10	ADC08455
19	19.2	76.8	2000	12	ADJ41650

20	19.2	76.8	6048	6	ABJ32509	ABJ32509 Human imm
21	19.2	76.8	6051	12	ADQ21014	Adq21014 Human sof
22	19.2	76.8	8333	4	AA545406	AA545406 Chemical1
23	19.2	76.8	8333	6	ABL33502	ABL33502 Human imm
24	19.2	76.8	8333	6	ABK28255	ABK28255 DNA trans
25	19.2	76.8	9483	6	ABJ32377	ABJ32377 Human imm
26	19.2	76.8	9483	6	ABJ70516	ABJ70516 Chemical1
27	19.2	76.8	9483	6	AA561092	AA561092 Human gen
28	19.2	76.8	13996	6	AA531523	AA531523 Human DNA
29	19.2	76.8	13996	6	ABQ66847	ABQ66847 Human pol
30	19.2	76.8	13996	10	ADCI1134	ADCI1134 Human DNA
31	19.2	76.8	14001	4	AA531522	AA531522 Human DNA
32	19.2	76.8	14001	6	ABQ66846	ABQ66846 Human pol
33	19.2	76.8	14001	10	ADCI1133	ADCI1133 Human DNA
34	19.2	76.8	26230	11	ACN44678	ACN44678 Human gen
35	19.2	76.8	28860	4	ABJ09806	ABJ09806 Drosoph11
36	19.2	76.8	32433	9	ADA02630	ADA02630 Human FLT
37	19.2	76.8	32433	10	ADB72368	ADB72368 Human FLT
38	19.2	76.8	32433	10	ADB95878	ADB95878 Human FLT
39	19.2	76.8	39401	14	ADZ13218	ADZ13218 Murine ca
40	19.2	76.8	43991	13	ABD32684	ABD32684 Human can
41	19.2	76.8	96597	10	ADC85287	ADC85287 Mouse fis
42	19.2	76.8	96597	9	ADA02807	ADA02807 Mouse fis
43	19.2	76.8	96597	10	ADB72545	ADB72545 Mouse fis
44	19.2	76.8	96597	12	ADM74402	ADM74402 Murine ca
45	19.2	76.8	110000	8	ABX16390_1	Continuation (2 of
46	19.2	76.8	117730	14	ADZ12550	Adz12550 Human can
47	19.2	76.8	196686	11	ACN44170	ACN44170 Human gen
48	19.2	76.8	197775	11	ACN44416	ACN44416 Mouse gen
49	18.8	75.2	41	12	ADN98686	ADN98686 Human 202
50	18.8	75.2	42	3	AA337946	AA337946 DNA synth
51	18.8	75.2	42	12	ADL19444	Adl19444 125P5C8 g
52	18.8	75.2	43	3	AA294401	AA294401 CDNA synth
53	18.8	75.2	43	3	AA293041	AA293041 Primer us
54	18.8	75.2	43	3	AA294895	AA294895 CDNA synth
55	18.8	75.2	43	3	AA14804	AA14804 Oligonuc1
56	18.8	75.2	43	3	AA09164	AA09164 CDNA synth
57	18.8	75.2	43	4	AD066225	AD066225 Human SGP
58	18.8	75.2	43	4	AD04804	Ad04804 Human sec
59	18.8	75.2	43	4	AA76005	AA76005 RT primer
60	18.8	75.2	43	6	ABJ50400	ABJ50400 Human 158
61	18.8	75.2	43	6	ABJ50412	ABJ50412 Human 158
62	18.8	75.2	43	6	ABN03602	ABN03602 Human CDN
63	18.8	75.2	43	6	AA599436	AA599436 DPNCN CD
64	18.8	75.2	43	6	ABK67415	ABK67415 Human 83P
65	18.8	75.2	43	6	AAJ53469	AAJ53469 zinc tran
66	18.8	75.2	43	8	ABZ78169	ABZ78169 CDNA synth
67	18.8	75.2	43	8	ABZ20556	ABZ20556 Cancer as
68	18.8	75.2	43	10	ADC71176	Adc71176 CDNA synth
69	18.8	75.2	43	10	ADD84526	Add84526 DPNCN CD
70	18.8	75.2	43	10	ABV99869	ABV99869 Human 121
71	18.8	75.2	43	11	ADJ97735	Adj97735 238P182 g
72	18.8	75.2	43	12	ADM83217	Adm83217 191P4D12(
73	18.8	75.2	43	12	ADM83820	Adm83820 Human can
74	18.8	75.2	43	12	ADN49850	Adn49850 238P182 c
75	18.8	75.2	43	12	ADQ64540	Adq64540 Human 213
76	18.8	75.2	44	12	ADQ68089	Adq68089 Cancer re
77	18.8	75.2	44	5	AA511655	AA511655 Prostate
78	18.8	75.2	107	3	AAA14809	AAA14809 Fragment
79	18.8	75.2	124	8	ABZ20567	Abz20567 Cancer as
80	18.8	75.2	192	8	ABZ78104	Abz78104 Human sup
81	18.8	75.2	192	13	ADU98925	Adu98925 Human 109
82	18.8	75.2	345	8	ABZ20564	Abz20564 Cancer as
83	18.8	75.2	426	5	AA511664	AA511664 84P2A9 su
84	18.8	75.2	686	6	ABQ29798	ABQ29798 Oligonuc1
85	18.8	75.2	686	6	ABQ29798	Abq29798 Oligonuc1
86	18.8	75.2	945	10	ADQ03876	Adq03876 Bacterial
87	18.8	75.2	1860	3	AA51651	AA51651 Arabidops
88	18.8	75.2	2000	6	ABZ14945	Abz14945 Arabidops
89	18.8	75.2	2033	3	AA50817	AA50817 Arabidops
90	18.8	75.2	2033	3	AA51556	AA51556 Arabidops
91	18.8	75.2	2037	3	AA51556	AA51556 Arabidops
92	18.8	75.2	138808	12	ADQ97928	Adq97928 Human can


```
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 9008.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR MPI; 2001-656860/75.
XX
DR P-PSDB; ABB60739..
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
PS Claim 1; SEQ ID NO 9008; 21bp + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4254 BP; 1253 A; 999 C; 1013 G; 989 T; 0 U; 0 Other;
XX
Query Match 80.8%; Score 20.2; DB 4; Length 4254;
Best Local Similarity 88.0%; Pred. No. 3.2e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAACTAAAGCTGATCTTC 25
DB 826 AAAAAAAAACTAAAGCTGATATTC 802
XX
RESULT 3
AB103756/C
ID ABL03756 standard; cDNA; 6746 BP.
XX
XX ABL03756;
XX
DT 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5750.
DE Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
XX
XX WO200171042-A2.
XX
XX
XX 27-SEP-2001.
XX
```

```
PF 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
DR MPI; 2001-656860/75.
XX
DR P-PSDB; ABB59653.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 5750; 21bp + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6746 BP; 1670 A; 1626 C; 1656 G; 1794 T; 0 U; 0 Other;
XX
Query Match 80.8%; Score 20.2; DB 4; Length 6746;
Best Local Similarity 88.0%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 1 AAAAAAAAACTAAAGCTGATCTTC 25
DB 5952 AAAAAAAAACTAAAGCTGATATTC 5928
XX
RESULT 4
AAS46788/C
ID AAS46788 standard; DNA; 61020 BP.
XX
XX AAS46788;
XX
DT 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #514.
DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP002955.
PF
XX
XX 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX MPI; 2001-602752/68.
DR
```

```
XX Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
XX Claim 1; SEQ ID NO 514; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumor suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analyzing
CC diseases associated with CpG dinucleotides e.g. cancers and tumors. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumor suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 4; Length 61020;
XX Best Local Similarity 88.0%; Pred. No. 3.6e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 25658 AAAAAAAAACTAAGCTTGATCTTC 25634
XX
XX RESULT 5
XX ID ADQ97611 standard; DNA; 62154 BP.
XX
XX ADQ97611;
XX
XX 07-OCT-2004 (first entry)
XX
XX Human cancer associated sequence HD10-014, SEQ ID 588.
XX
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
XX Homo sapiens.
XX
XX WO2004060304-A2.
XX
XX 22-JUL-2004.
XX
XX 22-DEC-2003; 2003WO-US041389.
XX
XX 27-DEC-2002; 2002US-00330773.
XX
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX
XX WPI; 2004-543781/52.
XX
XX New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
```

```
PT cancers such as leukemia and lymphoma.
XX
XX Claim 1; SEQ ID NO 588; 199pp; English.
XX
XX The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 62154 BP; 18095 A; 14107 C; 13141 G; 16811 T; 0 U; 0 Other;
XX
XX Query Match 80.8%; Score 20.2; DB 12; Length 62154;
XX Best Local Similarity 88.0%; Pred. No. 3.7e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 7984 AAAAAAAAACTAAGCTTGATCTTC 8008
XX
XX RESULT 6
XX ABD32968 3/c
XX Continuation (4 of 8) of ABD32968 from base 300001 (Human cancer-associated genomic DNA
XX Sequence split into 8 fragments LOCUS ABD32968 Accession Abd32968
XX Fragment Name Begin End
XX WP ABD32968_0 1 110000
XX WP ABD32968_1 100001 210000
XX WP ABD32968_2 200001 310000
XX WP ABD32968_3 300001 410000
XX WP ABD32968_4 400001 510000
XX WP ABD32968_5 500001 610000
XX WP ABD32968_6 600001 710000
XX WP ABD32968_7 700001 779603
XX
XX Query Match 80.8%; Score 20.2; DB 13; Length 110000;
XX Best Local Similarity 88.0%; Pred. No. 3.8e+02;
XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 18233 AAAAAAAAACTAAGCTTGATCTTC 18209
XX
XX RESULT 7
XX ACN45174/c
XX ID ACN45174 standard; DNA; 235070 BP.
XX
XX ACN45174;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human genomic sequence hCG15927.
XX
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
XX Homo sapiens.
XX
XX WO2003073826-A2.
XX
XX 12-SEP-2003.
XX
XX 28-FEB-2003; 2003WO-US006235.
XX
XX 01-MAR-2002; 2002US-00087192.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
XX
XX WPI; 2003-328604/31.
XX
XX
```

PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 1990; Opp; English.

XX The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published

XX
 SQ Sequence 235070 BP; 57319 A; 52466 C; 56014 G; 69049 T; 0 U; 222 Other;

Query Match 80.8%; Score 20.2; DB 11; Length 235070;
 Best Local Similarity 88.0%; Pred. No. 3.9e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAAGCTTGATCTTC 25
 |||||
 DB 126963 AAAAAAAAAAAGATGATCTTC 126939

RESULT 8
 AAS45458/c
 ID AAS45458 standard; DNA; 13606 BP.

XX
 AC AAS45458;
 XX

DT 18-DEC-2001 (first entry)

DE Chemically pretreated genomic DNA associated with cell cycle #82.

XX
 XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW gratic-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KW PCR primer.

XX
 OS Homo sapiens.
 XX

XX WO200168911-A2.
 XX

XX 20-SEP-2001.
 XX

XX 15-MAR-2001; 2001WO-EP002945.
 XX

XX 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX

XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2001-602751/68.
 DR

XX Designing primers and probes for analyzing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle.

PS Claim 1; SEQ ID NO 163; 28pp; English.

XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers

XX
 SQ Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;

Query Match 79.2%; Score 19.8; DB 4; Length 13606;
 Best Local Similarity 91.3%; Pred. No. 4.8e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAAGCTTGATCT 23
 |||||
 DB 5438 AAAAAAAAACTAAATGATCT 5416

RESULT 9
 AAS46562/c
 ID AAS46562 standard; DNA; 13606 BP.

XX
 AC AAS46562;
 XX

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #284.

XX
 XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
 KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.

XX
 OS Homo sapiens.
 XX

XX WO200168912-A2.
 XX

XX 20-SEP-2001.
 XX

XX 15-MAR-2001; 2001WO-EP002955.
 XX

XX 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX

XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2001-602752/68.
 DR

XX Fragments of chemically modified genes associated with tumor suppressor
 PT genes and oncogenes, useful in designing primers and probes for analyzing
 PT diseases associated with cytosine methylation state e.g. cancer.

XX
 PS Claim 1; SEQ ID NO 284; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and oncogenes
 CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
 CC 500 are missing from the sequence listing) sequences (Ss) and sequences

CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

CC Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;

Query Match 79.2%; Score 19.8; DB 4; Length 13606;
Best Local Similarity 91.3%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCT 23
Db 5438 AAAAAAAAACTAAAGCTTGATCT 5416

RESULT 10
ABL33811/C
ID ABL33811 standard; DNA; 13606 BP.

AC ABL33811;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1784.

XX Human; immune system diseases; cytosine methylation; antiasthmatic;
XX antileukemic; anti-HIV; anticonvulsant; ophthalmological;
XX neuroprotective; anti-rheumatic; antidiabetic; antipsychotic;
XX antirheumatic; antirheumatic; antidiabetic; antipsychotic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PE 02-JUL-2001; 2001WO-EP007537.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

PT WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

PS Claim 1; SEQ ID NO 1784; 32pp + Sequence Listing; German.

XX

CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention

XX Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 13606;
Best Local Similarity 91.3%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCT 23
Db 5438 AAAAAAAAACTAAAGCTTGATCT 5416

RESULT 11
ABK28314/C
ID ABK28314 standard; DNA; 13606 BP.

AC ABK28314;

DT 23-APR-2002 (first entry)

DE DNA transcription associated complementary genomic DNA #94.

XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
XX PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
XX single nucleotide polymorphism; adenostine deaminase deficiency; cancer;
XX viral infection; Sezary syndrome; haematological disorder; tuberculosis;
XX immunological disorder; Werner syndrome; developmental disorder;
XX psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
XX neurodegenerative disorder; Mardenburg syndrome; Niemann-Pick disease;
XX myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
XX angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
XX polyglutamine disorder; solid tumour.

OS Unidentified.

PN WO200192565-A2.

PD 06-DEC-2001.

PE 06-APR-2001; 2001WO-EP003973.

PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

PT WPI; 2002-090046/12.

PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.

PS Claim 1; SEQ ID NO 188; 32pp; English.

CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the

CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenoviral deaminase
CC deficiency, viral infection, retroviral infection, Searcy syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, porriasis, Kieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
CC disease, HSR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office

SO Sequence 13606 BP; 3084 A; 285 C; 3764 G; 6470 T; 0 U; 3 Other;

Query Match 79.2%; Score 19.8; DB 6; Length 13606;
Best Local Similarity 91.3%; Pred. No. 4.8e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAAGCTTGATCT 23
DB 5438 AAAAAAAAACTAAAGCTTGATCT 5416

RESULT 12
ABV07913/C
ID ABV07913 standard; cDNA; 387 BP.

AC ABV07913;
XX
XX
DT 13-SEP-2002 (first entry)
XX
XX
DE Human prostate expression marker cDNA 7904.

XX
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
XX
FN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183119P.
XX
PR 16-MAR-2000; 2000US-0189862P.
XX
PR 25-MAY-2000; 2000US-0207454P.
XX
PR 09-JUN-2000; 2000US-0211314P.
XX
PR 18-JUL-2000; 2000US-0219007P.
XX
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1, Page 1267; 11750pp; English.
XX
XX
CC The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (c) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker

XX
XX
Sequence 400 BP; 88 A; 97 C; 89 G; 116 T; 0 U; 10 Other;

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker

SO Sequence 387 BP; 92 A; 95 C; 77 G; 122 T; 0 U; 1 Other;

Query Match 76.8%; Score 19.2; DB 5; Length 387;
Best Local Similarity 87.5%; Pred. No. 6.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAAGCTTGATCT 24
DB 364 AAAAAAAAACTAAAGCTTGATCTT 341

RESULT 13
ABV07451/C
ID ABV07451 standard; cDNA; 400 BP.

AC ABV07451;
XX
XX
DT 13-SEP-2002 (first entry)
XX
XX
DE Human prostate expression marker cDNA 7442.

XX
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
XX
FN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183119P.
XX
PR 16-MAR-2000; 2000US-0189862P.
XX
PR 25-MAY-2000; 2000US-0207454P.
XX
PR 09-JUN-2000; 2000US-0211314P.
XX
PR 18-JUL-2000; 2000US-0219007P.
XX
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1, Page 1203; 11750pp; English.
XX
XX
CC The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (c) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker

XX
XX
Sequence 400 BP; 88 A; 97 C; 89 G; 116 T; 0 U; 10 Other;

Query Match 76.8%; Score 19.2; DB 5; Length 400;
Best Local Similarity 87.5%; Pred. No. 6.8e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTGATCTT 24
Db 398 AAAAAAAAACTAATCTTATCTT 375

RESULT 14

ABV04081/c
ID ABV04081 standard; cDNA; 424 BP.

XX
AC ABV04081;

XX
DT 13-SEP-2002 (first entry)

XX
DE Human prostate expression marker cDNA 4072.

XX
KW Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.

XX
OS Homo sapiens.

XX
PN WO200160860-A2.

XX
PD 23-AUG-2001.

XX
PF 20-FEB-2001; 2001WO-US005171.

XX
PR 17-FEB-2000; 2000US-0183319P.

XX
PR 16-MAR-2000; 2000US-0189862P.

XX
PR 25-MAY-2000; 2000US-0207454P.

XX
PR 09-JUN-2000; 2000US-0211314P.

XX
PR 18-JUL-2000; 2000US-0215007P.

XX
PR 13-DEC-2000; 2000US-0255281P.

XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX
PI Schlegel R, Endege WO, Monahan JB;

XX
DR WPI; 2001-662795/76.

XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
prostate cells and correlating with presence of prostate cancer, useful
for detecting presence of prostate cancer, stage of prostate cancer.

XX
PS Claim 1; Page 716; 11750pp; English.

XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification or its complement. (I) is useful for: (a) assessing whether
a patient is afflicted with prostate cancer; (b) monitoring the
progression of prostate cancer in a patient; (c) assessing the efficacy
of a test compound to inhibit prostate cancer in a patient; (d) assessing
the efficacy of a therapy for inhibiting prostate cancer in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the prostate cell carcinogenic potential of a compound; (g)
determining whether prostate cancer has metastasized in a patient; (h)
assessing the aggressiveness or indolence of prostate cancer in a patient
; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX
SQ Sequence 424 BP; 101 A; 99 C; 88 G; 133 T; 0 U; 3 Other;

Query Match 76.8%; Score 19.2; DB 5; Length 424;
Best Local Similarity 87.5%; Pred. No. 6.8e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTGATCTT 24
Db 373 AAAAAAAAACTAATCTTATCTT 350

RESULT 15
ADB87460/c
ID ADB87460 standard; DNA; 1834 BP.

XX
AC ADB87460;

XX
DT 04-DEC-2003 (first entry)

XX
DE Transgene expression regulatory element, STAR A21.

XX
KW gene transcription; regulatory; variety fragment; STAR element;
transgene expression; ds.

XX
OS Unidentified.

XX
PN WO2003004704-A2.

XX
PD 16-JAN-2003.

XX
PF 14-JUN-2002; 2002WO-NL000390.

XX
PR 04-JUL-2001; 2001EP-00202581.

XX
PR 05-JUL-2001; 2001US-0303199P.

XX
PA (CHRO-) CHROMAGENICS BV.

XX
PI Otte AP, Kruckeberg AL;

XX
DR WPI; 2003-229412/22.

XX
PT Selecting a DNA sequence with a gene transcription modulating quality by
providing a transcription system with a variety fragment-comprising
PT vectors and performing a selection step in the transcription system.

XX
PS Claim 43; Fig 26; 216pp; English.

XX
CC The invention relates to DNA sequences with gene transcription regulatory
sequences and methods for the detection and use of the regulatory DNA
sequences. The invention further comprises providing a transcription
CC system with a variety fragment-comprising vectors; and performing a
CC selection step in the transcription system in order to identify a
CC fragment comprising the DNA sequence with the gene transcription
CC modulating activity. This polynucleotide represents a STAR element used
in the method of the invention. The STAR element is shown to improve
transgene expression.

XX
SQ Sequence 1834 BP; 585 A; 298 C; 307 G; 644 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 10; Length 1834;
Best Local Similarity 87.5%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 AAAAAAAAACTAAGCTGATCTT 25
Db 791 AAAAAAAAACTAAGCTTATCATC 768

RESULT 16

ADJ35269/c
ID ADJ35269 standard; DNA; 1834 BP.

XX
AC ADJ35269;

XX
DT 22-APR-2004 (first entry)

XX
DE Thale cress stabilizing anti-repression, STAR, element #21.

XX
KW STAR affiliated proteinaceous molecule; post translational modification;
thale cress; stabilizing anti-repression; STAR; STAR element; ds.

XX
OS Arabidopsis thaliana.

CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence of a rice gene promoter. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpat_sequences.
CC
SQ Sequence 1999 BP; 497 A; 459 C; 428 G; 610 T; 0 U; 5 Other;
XX
Query Match 76.8%; Score 19.2; DB 10; Length 1999;
Best Local Similarity 87.5%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAACTAAGCTTATCTT 24
DB 695 AAAAAAAAAATAAGTTATTTT 672
RESULT 19
ADJ41650
ID ADJ41650 standard; cDNA; 2000 BP.
XX
AC ADJ41650;
XX
DT 06-MAY-2004 (first entry)
XX
DE Plant cDNA #2650.
XX
XX Plant; gene; ss; transcription; plant genome augmentation; cereal;
KM soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KM maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KM stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KM plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KM antifungal.
XX
OS Eukaryota.
XX
PN US2004016025-A1.
XX
PD 22-JAN-2004.
XX
PF 26-SEP-2002; 2002US-00260238.
XX
PR 26-SEP-2001; 2001US-0325277P.
XX 26-SEP-2001; 2001US-0325448P.
PR 04-APR-2002; 2002US-0370620P.
XX
PA (BUDW/) BUDWORTH P.
PA (MOUG/) MOUGHAMER T.
PA (BRIG/) BRIGGS S P.
PA (COOP/) COOPER B.
PA (GLAZ/) GLAZEBROOK J.
PA (GOFF/) GOFF S A.
PA (KATA/) KATAGIRI F.
PA (KREP/) KREPS J.
PA (PROV/) PROVART N.
PA (RICK/) RIQUE D.
PA (ZHUT/) ZHU T.
XX
PI Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI Goff SA, Katagiri F, Kreps J, Provart N, Rique D, Zhu T;
XX
DR WPI; 2004-190374/18.
XX
PT New rice promoter, useful for manipulating crop plants to alter or
PT improve phenotypic characteristics, e.g. produce large quantities of oil
PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT or high nutritional value.

XX
PS Claim 1; SEQ ID NO 2650; 230bp; English.
XX
XX The invention relates to plant nucleotide sequences that direct seed-,
CC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC or constitutive transcription of an operatively linked nucleic acid
CC segment. The invention also relates to a method for augmenting a plant
CC genome and a method of identifying a gene, where its expression is
CC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC encode are useful for manipulating crop plants to alter or improve
CC phenotypic characteristics, to produce large quantities of oil or
CC proteins, to incur resistance to insecticides, viruses or fungi, and to
CC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC have a high nutritional value with reduced apical dominance or dwarfism,
CC early flowering or altered metabolic pathways. This sequence represents a
CC plant nucleic acid of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification but was obtained in
CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
CC
SQ Sequence 2000 BP; 614 A; 369 C; 381 G; 628 T; 0 U; 8 Other;
XX
Query Match 76.8%; Score 19.2; DB 12; Length 2000;
Best Local Similarity 87.5%; Pred. No. 7.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAACTAAGCTTATCTT 24
DB 1344 AAAAAAAAAATAAGTTATTTT 1367
RESULT 20
ABL32509/C
ID ABL32509 standard; DNA; 6048 BP.
XX
AC ABL32509;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 482.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KM antiarteriosclerotic; antihaemic; cytoskeletal; noctropic;
KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KM antirheumatic; antiarthritic; antidiabetic; antiporiatic;
KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KM ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
PR
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 24

DB 524 AAAAAAAAACTAAAGCTTGATCTT 501

RESULT 23

ABL33502/C

ID ABL33502 standard; DNA; 8333 BP.

ABL33502;

26-MAR-2002 (first entry)

Human immune system associated gene SEQ ID NO: 1475.

Human; immune system disease; cytosine methylation; antiarthritic;
antiartherosclerotic; anti-HIV; anticonvulsant; ophthalmological;
neuroprotective; anti-rheumatic; antidiabetic; antipsychotic;
anti-inflammatory; cancer; eye disease; arteriosclerosis; anaemia;
acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.

Homo sapiens.

WO200200928-A2.

03-JAN-2002.

02-JUL-2001; 2001WO-EP007537.

30-JUN-2000; 2000DE-01032529.

01-SEP-2000; 2000DE-01043826.

(EPIC-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-130909/17.

Nucleic acid comprising fragment of chemically modified gene, useful for
diagnosis and treatment of diseases associated with abnormal cytosine
methylation.

Claim 1; SEQ ID NO 1475; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated
genes which are modified by the methylation of cytosines. The sequences
can be used in the diagnosis and treatment of immune system disorders,
including eye diseases such as retinopathy, neovascular glaucoma and
macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
diseases. The present sequence is a gene of the invention

Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 6; Length 8333;

Best Local Similarity 87.5%; Pred. No. 8e+02; Mismatches 3; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 24

DB 524 AAAAAAAAACTAAAGCTTGATCTT 501

RESULT 24

ABL33502/C

ID ABL33502 standard; DNA; 8333 BP.

AC ABL33502;

23-APR-2002 (first entry)

DNA transcription associated genomic DNA #65.

DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
PNA; cytosine methylation state; SNF; retroviral infection; gene; ds;
single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
viral infection; Sezary syndrome; haematological disorder; tuberculosis;
immunological disorder; Werner syndrome; developmental disorder;
psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
neurodegenerative disorder; Mardenburg syndrome; Niemann-Pick disease;
myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
polyglutamine disorder; solid tumour.

Unidentified.

WO200192565-A2.

06-DEC-2001.

06-APR-2001; 2001WO-EP003973.

06-APR-2000; 2000DE-01019058.

07-APR-2000; 2000DE-01019173.

30-JUN-2000; 2000DE-01032529.

01-SEP-2000; 2000DE-01043826.

(EPIC-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-090046/12.

New nucleic acid or oligomers, useful for diagnosing or treating
diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumours or
PT cancer.

Claim 1; SEQ ID NO 129; 32pp; English.

The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurodegenerative disorders, Mardenburg syndrome, Niemann-Pick disease,
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABL33502-ABL33502 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office

Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 76.8%; Score 19.2; DB 6; Length 8333;

Best Local Similarity 87.5%; Pred. No. 8e+02; Mismatches 3; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 24

DB 524 AAAAAAAAACTAAAGCTTGATCTT 501

RESULT 24

ABL33502/C

ID ABL33502 standard; DNA; 8333 BP.

DB	Sequence	Length	Score	DB	Length	Score
DB	524 AAAAAAAAACTAAACTTAACCTT 501	501				
RESULT 25						
ABL32377/c						
ID	ABL32377 standard; DNA; 9483 BP.					
XX	ABL32377;					
XX						
XX	26-MAR-2002 (first entry)					
DE	Human immune system associated gene SEQ ID NO: 350.					
XX						
KW	Human; immune system disease; cytosine methylation; antiasthmatic;					
KW	antiarteriosclerotic; antianaemic; cytosolic; nocrotropic;					
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;					
KW	antiinflammatory; antiarthritic; antidiabetic; antipsoriatic;					
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;					
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;					
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;					
KW	ds.					
XX						
OS	Homo sapiens.					
XX						
FN	WO200200928-A2.					
XX						
PD	03-JAN-2002.					
XX						
PF	02-JUL-2001; 2001WO-EP007537.					
XX						
PR	30-JUN-2000; 2000DE-01032529.					
PR	01-SEP-2000; 2000DE-01043826.					
PA	(EPIG-) EPIGENOMICS AG.					
XX						
F1	Olek A, Piepenbrock C, Berlin K;					
XX						
DR	WPI; 2002-130909/17.					
PT	Nucleic acid comprising fragment of chemically modified gene, useful for					
PT	diagnosis and treatment of diseases associated with abnormal cytosine					
XX	methylation.					
PS	Claim 1; SEQ ID NO 350; 32pp + Sequence Listing; German.					
XX						
CC	The present invention provides a number of human immune system associated					
CC	genes which are modified by the methylation of cytosines. The sequences					
CC	can be used in the diagnosis and treatment of immune system disorders,					
CC	including eye diseases such as retinopathy, neovascular glaucoma and					
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid					
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,					
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel					
XX	diseases. The present sequence is a gene of the invention					
XX						
SEQ	Sequence 9483 BP; 2366 A; 95 C; 1789 G; 4633 T; 0 U; 0 Other;					
Query Match	76.8%; Score 19.2; DB 6; Length 9483;					
Best Local Similarity	87.5%; Pred. No. 8e+02;					
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0.					
OY	1 AAAAAAAAACTAAAGCTTGATCTT 24					
DB	4181 AAAAAAAAAATAAACTTATCTT 4158					
RESULT 26						
ABL70516/c						
ID	ABL70516 standard; DNA; 9483 BP.					
XX	ABL70516;					
XX						
DT	01-JUL-2002 (first entry)					

XX	Chemically treated cell signalling DNA sequence complementary to#203.
DE	
KW	Cell signalling; cytosine methylation; cell signalling disease; cancer;
KM	tumour; cyostatic; ds.
XX	
OS	Unidentified.
XX	
PN	WO200202807-A2.
PD	
XX	
PD	10-JAN-2002.
XX	
PF	29-JUN-2001; 2001WO-EP007471.
XX	
PR	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
XX	
PA	(EPIC-) EPIDENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-154758/20.
XX	
PT	Nucleic acid, useful for diagnosis and therapy of diseases associated
PT	with cell signalling e.g. cancer, comprises chemically modified genomic
XX	sequences of genes associated with cell signalling.
PS	
XX	Claim 1; SEQ ID NO 406; 24pp + Sequence Listing; English.
CC	The invention relates to a nucleic acid comprising a sequence of at least
CC	18 bases of a segment of chemically pretreated DNA of genes associated
CC	with cell signalling. The activity of the modified sequences of the
CC	invention may be described as cyostatic. The object of the invention is
CC	to provide the chemically modified DNA of genes associated with cell
CC	signalling, as well as oligonucleotides and/or PNA-oligomers for
CC	detecting cytosine methylations, as well as a method which is
CC	particularly suitable for the diagnosis and/or therapy of genetic and
CC	epigenetic parameters of genes associated with cell signalling. The
CC	chemically modified DNA provided by the invention is useful for diagnosis
CC	and therapy of diseases such as solid tumours and cancer. The sequences
CC	given in records AB170111-AB170626 represent chemically pre-treated
CC	genomic DNA's of genes associated with cell signalling. Note: The
CC	sequence data for this patent is not represented in the printed
CC	specification, but is based on sequence information supplied by the
CC	European Patent Office
XX	
XX	Sequence 9483 BP; 2966 A; 95 C; 1789 G; 4633 T; 0 U; 0 Other;
XX	
Query Match	76.8%; Score 19.2; DB 6; Length 9483;
Best Local Similarity	87.5%; Pred. NO. 8e+02;
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0
QY	1 AAAAAAAAACTAAAGCTTGATCTT 24
DB	 4181 AAAAAAAAAATAAACTTATCTT 4158
RESULT 27	
AA561092/C	
ID	AA561092 standard; DNA; 9483 BP.
XX	
AC	AA561092;
XX	
DT	29-JAN-2002 (first entry)
XX	
DE	Human gene regulation-associated gene oligonucleotide #47.
KW	Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW	cardiac damage; inflammatory responses; Haemophilia; Werner syndrome;
KW	asthma; HBR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW	renal disease; Preeclampsia; cardiac allograft vascular disease;
KW	colorectal cancer; thyroid cancer; oesophageal cancer; de; tumour;
KW	immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;

KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cyostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003968.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
XX WPI: 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.
XX
PS Claim 1; SEQ ID NO 48; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, hemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, precociousness, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 9483 BP; 2966 A; 95 C; 1789 G; 4633 T; 0 U; 0 Other;
XX
Query Match 76.8%; Score 19.2; DB 6; Length 9483;
Best Local Similarity 87.5%; Pred. No. 8e+02; 3; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
CY 1 AAAAAAAAACTAAGCTGATCTT 24
Db 4181 AAAAAAAAAATAAAGCTTATCTT 4158
XX
RESULT 28
AAS31523
ID AAS31523 standard; DNA; 13996 BP.
XX
AC AAS31523;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human DNA for a novel extracellular matrix protein, Seq ID No 602.
XX
KW Human; secreted extracellular matrix protein; ds; immunomodulatory;
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; opthalmic; cytostatic;
KW antialzheimer; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;

KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.
XX
OS Homo sapiens.
XX
PN WO20015368-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001348.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0231415P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.

PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234423P.
PR 21-SEP-2000; 2000US-0234474P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235634P.
PR 27-SEP-2000; 2000US-0235636P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236602P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241212P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244674P.
PR 08-NOV-2000; 2000US-0244675P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246539P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465572/50.
DR
XX
PT Nucleic acid molecules encoding human secreted extracellular matrix
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers.
XX
XX Claim 1; SEQ ID NO 602; 577bp; English.
CC
CC The invention relates to isolated nucleic acid molecules encoding novel
CC human secreted extracellular matrix proteins (SPs). The polynucleotides
CC and proteins are used to prevent, treat a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For
CC example, disorders associated with decreased expression of SPs. The SP
CC polynucleotide or a vector expressing them may be administered to treat
CC diseases by gene therapy. Antisense molecules may be administered to down
CC regulate expression of SPs by binding with the cells own genes and
CC preventing their expression. The polynucleotides may also be used as DNA
CC probes in diagnostic assays. The SPs may also be used as antigens to
CC produce antibodies and to identify modulators (agonists and antagonists)
CC of the SPs. The anti-(SP) antibodies and antagonists may also be used to
CC down regulate expression and activity of SP and as diagnostic agents for
CC detecting the presence of SPs in samples. The disorders include for
CC example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency
CC virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis),
CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
CC diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-
CC /cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina),
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC (e.g. corneal infections). Other uses include wound healing, maintenance
CC of organs before transplantation, support of cell culture of primary
CC tissues, modulation of for example differentiation of embryonic stem

Query Match 76.8%; Score 19.2; DB 4; Length 13996;
Best Local Similarity 87.5%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 4306 AAAATTAACAAACCTGATCTT 4329

RESULT 29
AB066847
ID AB066847 standard; DNA; 13996 BP.
XX AB066847;
AC
XX
XX 23-AUG-2002 (first entry)
DT
XX
XX Human polynucleotide SEQ ID NO 602.
DE
XX
XX Human; noctropic; neuroprotective; cytoskeletal; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antislaking; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatocellular; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; anticancer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
OS
XX
XX US2002042386-A1.
PN
XX

PD 11-APR-2002. 2001US-00764870.
XX 17-JAN-2001; 2001US-00764870.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225470P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225575P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX
XX MPI, 2002-470713/50.
XX
XX New nucleic acid encoding human proteins, useful for diagnosis, treatment
XX and prevention of e.g. osteoporosis, also related polypeptides and
XX antibodies.
XX
XX Disclosure; SEQ ID NO 602; 235pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AB066521-AB066785) and proteins
XX (ABP47846-ABP48110) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.

CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. The present sequence is that of a polynucleotide of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?ocid=99999764870
XX
SQ Sequence 13996 BP; 4021 A; 3069 C; 2845 G; 4061 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 6; Length 13996;
Best Local Similarity 87.5%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTTAAGCTTGATCTT 24
Db 4306 AAAATPAAAAAATAAAGCTTGATCTT 4329
RESULT 30
ADCL1134
ID ADCL1134 standard; DNA; 13996 BP.
XX
XX AC ADCL1134;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE Human DNA from extracellular matrix gene 98 #2.
XX
XX KW Extracellular matrix protein; cytosolic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal; candidant;
KW cardiovascular; nephrotropic; antiinflammatory; musclic; Gen;
KW respiratory; immunosuppressive; cerebroprotective; vasotrophic;
KW neoplasic; anti-allergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; Human; ds.
XX
OS Homo sapiens.
XX
XX PN US2003059875-A1.
XX
XX PD 27-MAR-2003.
XX
XX PF 19-APR-2002; 2002US-00125540.
XX
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.

[illegible]

CC	PR	08-NOV-2000;	2000US-0246478P.
CC	PR	08-NOV-2000;	2000US-0246523P.
CC	PR	08-NOV-2000;	2000US-0246524P.
CC	PR	08-NOV-2000;	2000US-0246525P.
CC	PR	08-NOV-2000;	2000US-0246526P.
CC	PR	08-NOV-2000;	2000US-0246527P.
CC	PR	08-NOV-2000;	2000US-0246528P.
CC	PR	08-NOV-2000;	2000US-0246532P.
CC	PR	08-NOV-2000;	2000US-0246609P.
CC	PR	08-NOV-2000;	2000US-0246610P.
CC	PR	08-NOV-2000;	2000US-0246611P.
CC	PR	08-NOV-2000;	2000US-0246613P.
CC	PR	17-NOV-2000;	2000US-0249207P.
CC	PR	17-NOV-2000;	2000US-0249208P.
CC	PR	17-NOV-2000;	2000US-0249209P.
CC	PR	17-NOV-2000;	2000US-0249210P.
CC	PR	17-NOV-2000;	2000US-0249211P.
CC	PR	17-NOV-2000;	2000US-0249212P.
CC	PR	17-NOV-2000;	2000US-0249213P.
CC	PR	17-NOV-2000;	2000US-0249214P.
CC	PR	17-NOV-2000;	2000US-0249215P.
CC	PR	17-NOV-2000;	2000US-0249216P.
CC	PR	17-NOV-2000;	2000US-0249217P.
CC	PR	17-NOV-2000;	2000US-0249218P.
CC	PR	17-NOV-2000;	2000US-0249244P.
CC	PR	17-NOV-2000;	2000US-0249245P.
CC	PR	17-NOV-2000;	2000US-0249264P.
CC	PR	17-NOV-2000;	2000US-0249265P.
CC	PR	17-NOV-2000;	2000US-0249297P.
CC	PR	17-NOV-2000;	2000US-0249299P.
CC	PR	17-NOV-2000;	2000US-0249300P.
CC	PR	01-DEC-2000;	2000US-0250160P.
CC	PR	01-DEC-2000;	2000US-0250391P.
CC	PR	05-DEC-2000;	2000US-0251030P.
CC	PR	05-DEC-2000;	2000US-0251988P.
CC	PR	05-DEC-2000;	2000US-0256719P.
CC	PR	08-DEC-2000;	2000US-0251479P.
CC	PR	08-DEC-2000;	2000US-0251856P.
CC	PR	08-DEC-2000;	2000US-0251868P.
CC	PR	08-DEC-2000;	2000US-0251869P.
CC	PR	08-DEC-2000;	2000US-0251989P.
CC	PR	08-DEC-2000;	2000US-0251990P.
CC	PR	11-DEC-2000;	2000US-0254097P.
CC	PR	05-JAN-2001;	2001US-0259678P.
CC	PR	17-JAN-2001;	2001US-00764870.
CC	PA	(HUMA-)	HUMAN GENOME SCI INC.
CC	XX		
CC	XX		
CC	PI	Rosen CA, Ruben SM, Barash SC;	
CC	XX		
CC	XX	WPI; 2003-743765/70.	
CC	PT	New isolated nucleic acids and polypeptides, useful for diagnosing,	
CC	PT	treating, and/or preventing disorders, such as cancer, infections,	
CC	PT	cardiovascular and inflammatory diseases.	
CC	XX		
CC	PS	Disclosure; SEQ ID NO 602; 235bp; English.	
CC	XX		
CC	CC	The invention relates to an isolated nucleic acid molecule (cDNA)	
CC	CC	encoding a human extracellular matrix protein, representing one of 161	
CC	CC	novel genes. Also included are recombinant vectors, host cells	
CC	CC	(expressing the protein), the extracellular matrix proteins (including	
CC	CC	their fragments, epitopes and homologues), an isolated antibody that	
CC	CC	binds specifically to the protein, diagnosing a pathological condition or	
CC	CC	susceptibility to a pathological condition (comprising determining the	
CC	CC	presence or absence of a mutation in the nucleic acid and diagnosing a	
CC	CC	condition based on the presence or absence of the mutation), diagnosing a	
CC	CC	pathological condition or susceptibility to a pathological condition	
CC	CC	(comprising determining the presence or amount of expression of the	
CC	CC	protein in a biological sample and diagnosing a condition based on the	
CC	CC	presence or amount of expression of the protein), preventing, treating or	
CC	CC	ameliorating a medical condition by administering the nucleic acid or	
CC	CC	protein to a mammalian subject, identifying a binding partner to the	

CC protein, the gene corresponding to the cDNA sequence, and identifying an
CC activity in a biological assay (comprising expressing the nucleic acid in
CC a cell, isolating the supernatant, detecting an activity in a biological
CC assay and identifying the protein in the supernatant having the
CC activity). The nucleic acids and proteins display the following
CC activities: Cytostatic, antibacterial, Virucide, Neuroprotective,
CC Gynaecological, Gastrointestinal-Gen, Cardiant, Cardiovascular-Gen,
CC Nephrologic, Antiinflammatory, Muscular-Gen, Respiratory-Gen,
CC Immunosuppressive, Cerebroprotective, Vasotrophic, Nootropic,

Query Match 76.8%; Score 19.2; DB 10; Length 13996;
Best Local Similarity 87.5%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAACTGATCTT 24
Db 4306 AAAAAAACAACAACTGATCTT 4329

RESULT 31
AAS31522
ID AAS31522 standard; DNA, 14001 BP.
XX
AC AAS31522;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human DNA for a novel extracellular matrix protein. Seq ID No 601.
XX
KW Human; secreted extracellular matrix protein; ds; immunomodulatory;
KW Anti-HIV; antianemic; antineumatic; antisclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cyostatic;
KW antialzheimer; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; Rheumatoid arthritis; multiple sclerosis;
KW cancer; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.
XX
OS Homo sapiens.
XX
PN MO20015368-A1.
XX
BD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001348.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226661P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236328P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250191P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0254990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI, 2001-465572/50.
XX
XX Nucleic acid molecules encoding human secreted extracellular matrix
XX PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
XX PT Alzheimer's and Parkinson's diseases and cancers.
XX
XX
XX Claim 1; SEQ ID NO 601; 577bp; English.
XX
XX The invention relates to isolated nucleic acid molecules encoding novel
XX CC human secreted extracellular matrix proteins (SPs). The polynucleotides
XX CC and proteins are used to prevent, treat a medical condition in e.g.
XX CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. For
XX CC example, disorders associated with decreased expression of SPs. The SP
XX CC polynucleotide or a vector expressing them may be administered to treat
XX CC diseases by gene therapy. Antisense molecules may be administered to down
XX CC regulate expression of SPs by binding with the cells own genes and
XX CC preventing their expression. The polynucleotides may also be used as DNA
XX CC probes in diagnostic assays. The SPs may also be used as antigens to
XX CC produce antibodies and to identify modulators (agonists and antagonists)
XX CC of the SPs. The anti-(SP) antibodies and antagonists may also be used to
XX CC down regulate expression and activity of SP and as diagnostic agents for
XX CC detecting the presence of SPs in samples. The disorders include for
XX CC example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency
XX CC virus), hepatitis, anemia, rheumatoid arthritis and multiple sclerosis),
XX CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
XX CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
XX CC diseases (e.g. Alzheimer's disease, Parkinson's disease) cardio-
XX CC /cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina),
XX CC infections caused by bacteria, viruses and fungi and ocular disorders
XX CC (e.g. corneal infections). Other uses include wound healing, maintenance

CC of organs before transplantation, support of cell culture of primary
CC tissues, modulation of for example differentiation of embryonic stem
CC
Query Match 76.8%; Score 19.2; DB 4; Length 14001;
Best Local Similarity 87.5%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTT 24
DB 4307 AAAAAATTAACAAAACCTTGATCTT 4330
RESULT 32
AB066846
ID AB066846 strand; DNA; 14001 BP.
XX
AC AB066846;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 601.
XX
XX Human; nocrotropic; neuroprotective; cytosratic; dermatological; vitricide;
XX KW immunosuppressive; anti-HIV; antibacterial; vulniterary;
XX KW antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;
XX KW antirheumatic; hepatocrotropic; cerebroprotective; antiinflammatory;
XX KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
XX KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS Homo sapiens.
XX
XX
XX US2002042386-A1.
XX
PD 11-APR-2002.
XX
XX
PF 17-JAN-2001; 2001US-00764870.
XX
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236386P.
PR 29-SEP-2000; 2000US-0236393P.

PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
PI WPI; 2002-470713/50.
XX
XX New nucleic acid encoding human proteins, useful for diagnosis, treatment
PT and prevention of e.g. osteoporosis, also related polypeptides and
PT antibodies.
XX
XX Disclosure; SEQ ID NO 601; 235bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AB065521-AB066785) and proteins
CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune chryoiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. The present sequence is that of a polynucleotide of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?docID=99909764870
XX
XX Sequence 14001 BP; 4025 A; 3074 C; 2841 G; 4061 T; 0 U; 0 Other;
SQ

Query Match 76.8%; Score 19.2; DB 6; Length 14001;
Best Local Similarity 87.5%; Pred. No. 8.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 4307 AAAATAAACAAAACTTGATCTT 4330

RESULT 33
ADCL1133
ID ADCL1133 standard; DNA; 14001 BP.
XX
XX ADCL1133;
AC
XX
XX 18-DEC-2003 (first entry)
XX
XX Human DNA from extracellular matrix gene 98 #1.
XX
XX Extracellular matrix protein; cytosolic; antibacterial; virucide;
KM neuroprotective; gynaecological; gastrointestinal; cardiac;
KM cardiovascular; Gen; nephrotropic; antiinflammatory; muscular; Gen;

KM respiratory; Gen; immunosuppressive; cerebroprotective; vasotropic;
KM neutrotropic; anti-allergic; cancer; bacterial infection; viral infection;
KM neutral disorder; immune system disorder; blood disorder;
KM muscular disorder; reproductive disorder; gastrointestinal disorder;
KM pulmonary disorder; cardiovascular disorder; renal disorder;
KM inflammatory disorder; proliferative disorder; Human; ds.
OS Homo sapiens.
XX
XX US2003059875-A1.
XX
XX 27-MAR-2003.
XX
XX 19-APR-2002; 2002US-00125540.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225265P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226686P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0232081P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 05-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764870.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX PI Rosen CA, Ruben SM, Barash SC;
XX
XX MPI; 2003-743765/70.
XX
XX
XX PT New isolated nucleic acids and polypeptides, useful for diagnosing,
XX PT treating, and/or preventing disorders, such as cancer, infections,
XX PT cardiovascular and inflammatory diseases.
XX
XX PS Disclosure; SEQ ID NO 601; 235pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule (cDNA)
XX CC encoding a human extracellular matrix protein, representing one of 161
XX CC novel genes. Also included are recombinant vectors, host cells
XX CC (expressing the protein), the extracellular matrix proteins (including
XX CC their fragments, epitopes and homologues), an isolated antibody that
XX CC binds specifically to the protein, diagnosing a pathological condition or
XX CC susceptibility to a pathological condition (comprising determining the
XX CC presence or absence of a mutation in the nucleic acid and diagnosing a
XX CC condition based on the presence or absence of the mutation), diagnosing a
XX CC pathological condition or susceptibility to a pathological condition
XX CC (comprising determining the presence or amount of expression of the
XX CC protein in a biological sample and diagnosing a condition based on the
XX CC presence or amount of expression of the protein), preventing, treating or
XX CC ameliorating a medical condition by administering the nucleic acid or
XX CC protein to a mammalian subject, identifying a binding partner to the
XX CC protein, the gene corresponding to the cDNA sequence, and identifying an
XX CC activity in a biological assay (comprising expressing the nucleic acid in
XX CC a cell, isolating the supernatant, detecting an activity in a biological
XX CC assay and identifying the protein in the supernatant having the
XX CC activity). The nucleic acids and proteins display the following
XX CC activities: Cytostatic, antibacterial, Virucide, Neuroprotective,
XX CC Gynaecological, Gastrointestinal-Gen, Cardiant, Cardiovascular-Gen,
XX CC Nephrologic, Antiinflammatory, Muscular-Gen, Respiratory-Gen,
XX CC Immunosuppressive, Cerebroprotective, Vasootropic, Neurotropic,
XX
XX
XX Query Match 76.8%; Score 19.2; DB 10; Length 14001;
XX Best Local Similarity 87.5%; Pred. No. 8.2e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAAAAAAAACTAAGCTTGATCTT 24
DB 4307 AAAAAATMAAACAACTTGATCTT 4330
RESULT 34
ACN44678/C
ID ACN44678 standard; DNA; 26230 BP.
XX
XX AC ACN44678;
XX
XX DT 18-NOV-2004 (first entry)
XX
XX DE Human genomic sequence hCG1737521.
XX
XX KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO2003073826-A2.
XX
XX PD 12-SEP-2003.
XX
XX PF 28-FEB-2003; 2003WO-US006235.
XX

PR 01-MAR-2002; 2002US-00087192.
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW;
XX WPI; 2003 328604/31.
XX
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
PR
XX
XX Claim 1; SEQ ID NO 1246; 0pp; English.
PS
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a bioclip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182566A1, for which no sequence data was published
XX
SQ Sequence 26230 BP; 7368 A; 4959 C; 5216 G; 8570 T; 0 U; 117 Other;
Query Match 76.8%; Score 19.2; DB 11; Length 26230;
Best Local Similarity 87.5%; Pred. No. 8.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTTAAGCTTGATCTT 24
Db 17550 AAAAAAAAAAGAAATCTTGATCTT 17527
RESULT 35
ABLO9806
ID ABLO9806 standard; cDNA; 28860 BP.
XX
XX ABL09806;
AC
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23900.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US009231.
PF
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 23-MAR-2000; 2000US 00614150.
XX
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers BW;
PI
XX WPI; 2001-656860/75.
DR
XX P-PSDB; ABB65703.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
PT

XX
XX Claim 1; SEQ ID NO 23900; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcr_sequences
XX
SQ Sequence 28860 BP; 8636 A; 5746 C; 5581 G; 8897 T; 0 U; 0 Other;
Query Match 76.8%; Score 19.2; DB 4; Length 28860;
Best Local Similarity 87.5%; Pred. No. 8.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTTAAGCTTGATCTT 24
Db 7230 AAAAAAAAACTCAAGCTTGTTT 7253
RESULT 36
ADA02630/C
ID ADA02630 standard; DNA; 32433 BP.
XX
XX ADA02630;
AC
XX
DT 06-NOV-2003 (first entry)
XX
DE Human FLT3 carcinoma associated gene, SEQ ID NO:1148.
XX
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
XX Homo sapiens.
OS
XX
XX WO2003057146-A2.
PN
XX
XX 17-JUL-2003.
PD
XX
XX 26-DEC-2002; 2002WO-US041414.
PF
XX
XX 26-DEC-2001; 2001US-00035832.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
XX Morris DW;
PI
XX WPI; 2003-587068/55.
DR
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1148; 245pp; English.
PS
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a bioclip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or

CC leukemia) or a propensity to carcinoma by determination of the sequence
 CC of a CA gene, or by determination of CA gene expression in particular
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as
 CC therapeutic agents and in screening and evaluating drug candidates. The
 CC present sequence represents a specifically claimed human CA nucleic acid
 CC sequence of the invention. Note: The complete sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcc_sequences.

XX Sequence 32433 BP; 8463 A; 7444 C; 7079 G; 9447 T; 0 U; 0 Other;

XX Query Match 76.8%; Score 19.2; DB 9; Length 32433;

XX Best Local Similarity 87.5%; Pred. No. 8.6e+02; Mismatches 3; Indels 0; Gaps 0;

XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX 1 AAAAAAAAACTAAAGCTTGATCTT 24

XX 1432 AAAAAAAAAAAAGCTTGATATT 1409

XX RESULT 37

XX ADB72368/c

XX ID ADB72368 standard; DNA; 32433 BP.

XX AC ADB72368;

XX DT 04-DEC-2003 (first entry)

XX DE Human FLT3 gene.

XX DE human; ds; cytosstatic; gene therapy; vaccine; carcinoma; lymphomas;

XX KM cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX OS Homo sapiens.

XX PN WO2003008583-A2.

XX PD 30-JAN-2003.

XX PF 26-DEC-2001; 2001WO-US051291.

XX PR 02-MAR-2001; 2001US-00798586.

XX PR 23-OCT-2001; 2001US-00004113.

XX PR 08-NOV-2001; 2001US-00052482.

XX PR 30-NOV-2001; 2001US-00997722.

XX PR 20-DEC-2001; 2001US-00034650.

XX PA (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX DR WPI; 2003-239337/23.

XX PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,

XX PT cancer; neoplasm, adenocarcinoma, or sarcomas.

XX PS Claim 1; SEQ ID NO 196; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a

XX nucleotide sequence selected from any of the 660 sequences fully defined

XX in the specification. A polynucleotide of the invention has cytosstatic

XX activity, and may have a use in gene therapy, or in a vaccine. The

XX recombinant nucleic acids and polypeptides are useful for treating

XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and

XX sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 32433 BP; 8463 A; 7444 C; 7079 G; 9447 T; 0 U; 0 Other;

XX Query Match 76.8%; Score 19.2; DB 10; Length 32433;

XX Best Local Similarity 87.5%; Pred. No. 8.6e+02; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTAAAGCTTGATCTT 24

Db 1432 AAAAAAAAAAAAGCTTGATATT 1409

RESULT 38

ADBE95878/c

ID ADB95878 standard; DNA; 32433 BP.

AC ADB95878;

DT 12-FEB-2004 (first entry)

DE Human FLT3 gene genomic DNA sequence.

KW cancer diagnosis; cancer treatment; carcinoma; cytosstatic; gene therapy;

KW lymphoma; breast cancer; prostate cancer; leukemia; ds; human; FLT3.

OS Homo sapiens.

PN WO2003039484-A2.

PD 15-MAY-2003.

PF 08-NOV-2002; 2002WO-US036071.

PR 08-NOV-2001; 2001US-00052482.

PA (SAGR-) SAGRES DISCOVERY.

PI Morris DW, Engelhard EK;

DR WPI; 2003-441462/41.

PT New carcinoma associated nucleic acids and proteins, useful for screening

PT drug candidates, or for diagnosing and treating carcinomas, e.g.

PT lymphoma, breast cancer, prostate cancer or leukemia.

PS Claim 1; SEQ ID NO 136; 793pp; English.

XX This invention relates to novel recombinant nucleic acids for use in

XX diagnosis and treatment of cancer, especially carcinomas, as well as the

XX use of compositions in screening methods. The compositions of the

XX invention may have cytosstatic activity whilst the disclosed sequences may

XX be useful for gene therapy. The carcinoma associated nucleic acids and

XX proteins are useful for diagnosing and treating carcinomas, for example

XX lymphoma, breast cancer, prostate cancer or leukemia, or for screening

XX drug candidates or bioactive agents capable of binding to, or modulating

XX the activity of, a carcinoma associated protein. The present sequence is

XX the genomic DNA sequence of the human FLT3 gene which is a carcinoma

XX associated gene of the invention.

XX Sequence 32433 BP; 8463 A; 7444 C; 7079 G; 9447 T; 0 U; 0 Other;

XX Query Match 76.8%; Score 19.2; DB 10; Length 32433;

XX Best Local Similarity 87.5%; Pred. No. 8.6e+02; Mismatches 3; Indels 0; Gaps 0;

XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX 1 AAAAAAAAACTAAAGCTTGATCTT 24

Db 1432 AAAAAAAAAAAAGCTTGATATT 1409

RESULT 39

ADZ13218/c

ID ADZ13218 standard; DNA; 39401 BP.

AC ADZ13218;

DT 16-JUN-2005 (first entry)

DE Murine cancer-associated genomic DNA #61.

OS Homo sapiens.
 PN WO2004074320-A2.
 XX
 PD 02-SEP-2004.
 XX
 PF 17-FEB-2004; 2004WO-US004730.
 XX
 PR 14-FEB-2003; 2003US-00367094.
 PR 14-MAR-2003; 2003US-00388838.
 PR 15-APR-2003; 2003US-00417375.
 PR 13-JUN-2003; 2003US-00461862.
 PR 15-SEP-2003; 2003US-00663431.
 PR 15-DEC-2003; 2003US-00737318.
 XX
 PA (SAGR-) SAGRES DISCOVERY INC.
 XX
 PI Morris DW, Morris DW, Malandro MS;
 XX
 DR WPI, 2004-652914/63.
 XX
 PT New isolated cancer-associated polynucleotides and polypeptides useful
 PT for diagnosing, preventing or treating cancers, especially lymphoma and
 PT leukemia, or in screening for agents that modulate cancer.
 XX
 PS claim 16; seqid 239; 310pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising at least 10
 CC contiguous nucleotides of any of the 233 polynucleotide sequences given
 CC in the specification, or its complement. The nucleic acids encode cancer-
 CC associated proteins. Also included are an expression vector comprising
 CC the isolated nucleic acid cited above, a host cell comprising the above
 CC recombinant nucleic acid or expression vector, a microarray for detecting
 CC a cancer-associated (CA) nucleic acid comprising at least one probe
 CC comprising at least 10 contiguous nucleotides of any of the above-
 CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
 CC an open reading frame of a CA sequence selected from any of the 95
 CC polynucleotide sequences as mentioned in the specification, or its
 CC complement), an isolated antibody, (or its antigen binding fragment) that
 CC binds to the above polypeptide, a hybridoma that produces the above
 CC monoclonal antibody, a pharmaceutical composition comprising the above
 CC antibody and a pharmaceutical excipient, a kit for detecting cancer
 CC cells (comprising the antibody cited above, methods for diagnosing cancer
 CC or for detecting the presence or absence of cancer cells in an
 CC individual, a method for inhibiting growth of cancer cells in an
 CC individual, a method for delivering a therapeutic agent to cancer cells
 CC in an individual, an electronic library comprising the above
 CC polynucleotide or polypeptide (or their fragments), methods of screening
 CC for anticancer activity or for a bioactive agent capable of modulating
 CC the activity of a CA protein (CAP), methods for detecting cancer
 CC associated with expression of a polypeptide in a test cell sample, a
 CC method for treating cancers and a method for inhibiting the expression of
 CC CA gene in a cell. The composition and methods are useful for detecting,
 CC diagnosing, preventing and treating cancers, especially lymphoma and
 CC leukemia. These may also be used in screening for agents that modulate
 CC cancer. The present sequence is a human CAP genomic sequence. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 41991 BP; 12474 A; 8216 C; 8278 G; 13023 T; 0 U; 0 Other;
 QY
 Query Match 76.8%; Score 19.2; DB 13; Length 41991;
 Best Local Similarity 87.5%; Pred. No. 8.7e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 1 AAAAAAAAACTAAGCTTGACTT 24
 ||||||| ||||||| |||||
 Db 35368 AAAAAAAAAATTAAAGCATCATCTT 35391

```

ID ADC85287 standard; DNA; 96596 BP.
XX
XX ADC85287;
AC
XX
XX 01-JAN-2004 (first entry)
DT
XX
XX Mouse Fish genomic sequence.
DE
XX
XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX
XX Mus sp.
OS
XX
XX WO2003045230-A2.
PN
XX
XX 05-JUN-2003.
PD
XX
XX 02-DEC-2002; 2002WO-US038582.
PF
XX
XX 30-NOV-2001; 2001US-00977722.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
XX Morris DW; Engelhard EK;
PI
XX
XX MPI; 2003-513603/48.
DR
XX
XX New recombinant nucleic acid comprising a nucleotide sequence of any of
PT the carcinoma-associated (CA) genes, useful for screening for drug
PT candidates for diagnosing or treating carcinomas.
XX
XX Claim 1; SEQ ID NO 73; 983bp; English.
XX
XX The invention relates to a recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the fully defined carcinoma-
CC associated (CA) genes from the 50 tables given in the specification. The
CC CA proteins are secreted, transmembrane or intracellular proteins. The
CC recombinant nucleic acids are useful for screening for drug candidates
CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC ADC85514 represent CA genes of the invention.
CC
SQ Sequence 96596 BP; 22664 A; 22752 C; 23978 G; 24296 T; 0 U; 2906 Other;

Query Match          76.8%; Score 19.2; DB 10; Length 96596;
Best Local Similarity 87.5%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAAGCTTGATCTTC 25
Db 4094 AAAAAAAAACTAAAGCTTGATCTTC 4117

RESULT 42
ADAO2807
ID ADAO2807 standard; DNA; 96597 BP.
XX
XX ADAO2807;
AC
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Mouse Fish carcinoma associated gene, SEQ ID NO:1325.
DE
XX
XX Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
XX Mus sp.
OS
XX
XX WO2003057146-A2.
PN
XX
XX 17-JUL-2003.
PD
XX
XX 26-DEC-2002; 2002WO-US041414.
PF

```

```

XX
XX 26-DEC-2001; 2001US-0005832.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
XX Morris DW;
PI
XX
XX MPI; 2003-587068/55.
DR
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
PT
XX
XX Claim 1; SEQ ID NO 1325; 245bp; English.
PS
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADAO1482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed murine CA nucleic acid
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 96597 BP; 22665 A; 22752 C; 23978 G; 24296 T; 0 U; 2906 Other;

Query Match          76.8%; Score 19.2; DB 9; Length 96597;
Best Local Similarity 87.5%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAAGCTTGATCTTC 25
Db 4094 AAAAAAAAACTAAAGCTTGATCTTC 4117

RESULT 43
ADB72545
ID ADB72545 standard; DNA; 96597 BP.
XX
XX ADB72545;
AC
XX
XX 04-DEC-2003 (first entry)
DT
XX
XX Mouse Fish gene.
DE
XX
XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX Mus sp.
OS
XX
XX WO2003008583-A2.
PN
XX
XX 30-JAN-2003.
PD
XX
XX 26-DEC-2001; 2001WO-US051291.
PF
XX
XX 02-MAR-2001; 2001US-00798586.
PR
XX
XX 23-OCT-2001; 2001US-00004113.
PR
XX
XX 08-NOV-2001; 2001US-00052482.
PR
XX
XX 30-NOV-2001; 2001US-00977722.
PF

```

```
PR 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW, Engelhard EK;
XX WPI; 2003-239337/23.
DR
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasms, adenocarcinoma, or sarcomas.
PT
XX Claim 1; SEQ ID NO 373; 2304pp; English.
PS
XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasms, adenocarcinoma, and
CC sarcomas. The present sequence represents a mouse gene of the invention.
XX
SQ Sequence 96597 BP; 22665 A; 22752 C; 23978 G; 24296 T; 0 U; 2906 Other;
Query Match 76.8%; Score 19.2; DB 10; Length 96597;
Best Local Similarity 87.5%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
DB 4094 AAAAAAAAACTAAGCTTGATCTTC 4117

RESULT 44
ADM74402
ID ADM74402 standard; DNA; 96597 BP.
XX
XX ADM74402;
XX
XX 01-JUL-2004 (first entry)
XX
XX Murine carcinoma associated (CA) nucleic acid #37.
XX
XX Mouse; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
XX carcinoma associated protein; CAP; carcinoma; leukemia; lymphoma;
XX cytostatic.
XX
XX Mus musculus.
XX
XX US2004072154-A1.
XX
XX 15-APR-2004.
XX
XX 30-NOV-2001; 2001US-00997722.
XX
XX 22-DEC-2000; 2000US-00747377.
XX 02-MAR-2001; 2001US-00798586.
XX
XX (MORR/) MORRIS D W.
XX (ENGE/) ENGELHARD E K.
XX
XX Morris DW, Engelhard EK;
XX WPI; 2004-328562/30.
XX
XX New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
XX Claim 1; SEQ ID NO 73; 29pp; English.
XX
XX The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
```

```
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukemia or lymphoma. This sequence represents a murine carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 96597 BP; 22665 A; 22752 C; 23978 G; 24296 T; 0 U; 2906 Other;
Query Match 76.8%; Score 19.2; DB 12; Length 96597;
Best Local Similarity 87.5%; Pred. No. 9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
DB 4094 AAAAAAAAACTAAGCTTGATCTTC 4117

RESULT 45
ABX16390_1
Continuation (2 of 7) of ABX16390 from base 100001 (Mouse high growth region.)
WP Sequence split into 7 fragments LOCUS ABX16390 Accession Abx16390
WP Fragment Name Begin End
WP ABX16390_0 1 110000
WP ABX16390_1 100001 210000
WP ABX16390_2 200001 310000
WP ABX16390_3 300001 410000
WP ABX16390_4 400001 510000
WP ABX16390_5 500001 610000
WP ABX16390_6 600001 659158

Query Match 76.8%; Score 19.2; DB 8; Length 110000;
Best Local Similarity 87.5%; Pred. No. 9.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTT 24
DB 56543 AAAAAAAAACTAAGCTTGATCTT 56566

RESULT 46
ADZ12550/C
ID ADZ12550 standard; DNA; 117730 BP.
XX
XX ADZ12550;
XX
XX 16-JUN-2005 (first entry)
XX
XX Human cancer-associated genomic DNA #8.
XX
XX
```



```
XX XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KM cytostatic; gene; ds.
XX OS Homo sapiens.
XX PN WO2005031001-A2.
XX PD 07-APR-2005.
XX PF 23-SEP-2004; 2004WO-US031617.
XX PR 23-SEP-2003; 2003US-00669920.
XX PA (CHIR ) CHIRON CORP.
XX PI Morris DW, Malandro MS;
XX DR WPI; 2005-273395/28.
XX PT Nucleic acid array useful for detecting cancer associated nucleic acid,
XX PS comprises two or more nucleic acid probes.
XX PS Disclosure; SEQ ID NO 70; 198pp; English.
XX CC The invention relates to a nucleic acid array for detecting a cancer
XX CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.
XX CC The invention also relates to a peptide array comprising two or more
XX CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound
XX CC that binds to a polypeptide, an isolated antibody or its fragment which
XX CC binds to a polypeptide, which is prepared by immunizing a host animal
XX CC with a composition comprising the polypeptide or its antigen binding
XX CC fragment and collecting cells from the host expressing antibodies against
XX CC the antigen or its antigen binding fragment, a composition comprising the
XX CC antibody and a carrier, a method of screening for anticancer activity, a
XX CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a
XX CC method of treating cancer and a method of inhibiting expression of a CA
XX CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
XX CC nucleic acids. The antibody is useful for detecting the presence or
XX CC absence of cancer cells in an individual which involves contacting cells
XX CC from the individual with the antibody and detecting a complex of a CA
XX CC protein from the cancer cells and the antibody, where the detection of
XX CC the complex correlates with the presence of cancer cells in the
XX CC individual. The composition is useful for inhibiting growth of cancer
XX CC cells in an individual or for delivering a therapeutic agent to cancer
XX CC cells in an individual. The invention is also useful for diagnosing
XX CC cancer, for treating cancer and for inhibiting expression of a CA gene in
XX CC a cell. This sequence represents human cancer-associated genomic DNA of
XX CC the invention.
XX SQ Sequence 117730 BP; 32629 A; 25620 C; 25334 G; 34147 T; 0 U; 0 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 14; Length 117730;
XX Best Local Similarity 87.5%; Pred. No. 9.1e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAACTAAAGCTTGATCTT 24
XX Db 65143 AAAAAAAAAAAAGCTTGATATT 65120
XX
XX RESULT 47
XX ACN44170/c
XX ID ACN44170 standard; DNA; 196686 BP.
XX AC ACN44170;
XX DT 18-NOV-2004 (first entry)
XX DE Human genomic sequence hCG39530.
XX XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX KM
```

```
OS Homo sapiens.
XX PN WO2003073826-A2.
XX PD 12-SEP-2003.
XX PF 26-FEB-2003; 2003WO-US006235.
XX PR 01-MAR-2002; 2002US-00087192.
XX PA (SAGR-) SAGRES DISCOVERY.
XX PI Morris DW;
XX DR WPI; 2003-328604/31.
XX PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX PS comprises a nucleotide sequence.
XX PS Claim 1, SEQ ID NO 484; 0pp; English.
XX CC The present invention relates to novel DNA and protein sequences which
XX CC are associated with carcinomas. The sequences are useful for: (i) for
XX CC screening drug candidates; (ii) for screening of bioactive agent capable
XX CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX CC a bioactive agent capable of modulating the activity of CAP; (iv) for
XX CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX CC determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX CC carcinoma including lymphoma. The present sequence is one such CA coding
XX CC sequence. Note: This patent is an equivalent to basic patent
XX US200218256A1, for which no sequence data was published
XX
XX SQ Sequence 196686 BP; 53978 A; 42758 C; 43862 G; 55372 T; 0 U; 716 Other;
XX
XX Query Match 76.8%; Score 19.2; DB 11; Length 196686;
XX Best Local Similarity 87.5%; Pred. No. 9.3e+02;
XX Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAACTAAAGCTTGATCTT 24
XX Db 93053 AAAAAATAAACAAAGCTTGATCTT 93030
XX
XX RESULT 48
XX ACN44416
XX ID ACN44416 standard; DNA; 197775 BP.
XX AC ACN44416;
XX DT 18-NOV-2004 (first entry)
XX DE Mouse genomic sequence mCG189928.
XX XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX KM Mus musculus.
XX OS
XX PN WO2003073826-A2.
XX PD 12-SEP-2003.
XX PF 26-FEB-2003; 2003WO-US006235.
XX PR 01-MAR-2002; 2002US-00087192.
XX PA (SAGR-) SAGRES DISCOVERY.
XX PI Morris DW;
XX DR
XX PT
XX PS
```

DR WPI; 2003-328604/31.
 XX
 PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 XX comprises a nucleotide sequence.
 XX
 PS Claim 1; SEQ ID NO 853; Opp; English.
 CC The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biobip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published
 CC
 SQ Sequence 197775 BP; 53733 A; 41437 C; 42826 G; 57644 T; 0 U; 2135 Other;
 Query Match 76.8%; Score 19.2; DB 11; Length 197775;
 Best Local Similarity 87.5%; Pred. No. 9.3e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 AAAAAAAAACTAAAGCTTGATCTTC 25
 Db 118648 AAAAAAAAACAAAAGCGTGATTTTC 118671
 RESULT 49
 ADN89686/C
 ID ADN89686 standard; DNA; 41 BP.
 XX
 AC ADN89686;
 XX
 DT 01-JUN-2004 (first entry)
 DE Human 202P5A5 cDNA synthesis primer.
 XX
 KW 202P5A5; human; cancer; tumour; ss; primer.
 XX
 OS Homo sapiens.
 XX
 FN WO2004016736-A2.
 PD 26-FEB-2004.
 PF 16-JUN-2003; 2003WO-US018906.
 PR 16-AUG-2002; 2002US-0404306P.
 PR 01-NOV-2002; 2002US-0423290P.
 XX
 PA (AGEN-) AGENSYS INC.
 XX
 PI Raitano AB, Paris M, Challita-Eid PM, Jakobovits A, Ge W;
 DR WPI; 2004-203774/19.
 XX
 PT New compositions having the 202P5A5 gene and encoded protein, useful for
 PT diagnosing, preventing, prognosticating or treating cancer of the
 PT prostate, bladder, colon, lung, ovary, breast, stomach, cervix, lymphoma,
 PT bone and/or skin.
 XX
 PS Example 1; SEQ ID NO 28; 266pp; English.
 CC The invention relates to a composition comprising 202P5A5 proteins. The
 CC composition and proteins are useful for detecting and treating cancer by
 CC inhibiting the growth or viability of cancer cells. The present sequence
 CC represents the human 202P5A5 cDNA synthesis primer.

XX
 SQ Sequence 41 BP; 3 A; 2 C; 2 G; 34 T; 0 U; 0 Other;
 Query Match 75.2%; Score 18.8; DB 12; Length 41;
 Best Local Similarity 90.9%; Pred. No. 8.6e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAAAGCTTGATC 22
 Db 26 AAAAAAAAAAAAGCTTGATC 5
 RESULT 50
 ID AAA37946/C
 ID AAA37946 standard; DNA; 42 BP.
 AC AAA37946;
 XX
 DT 18-AUG-2000 (first entry)
 DE DNA synthesis primer used in PTAN gene isolation.
 KW PTAN; testis specific; prostate cancer; overexpress; chromosome 1q22;
 KW diagnose; cancer; breast; vaccine; primer; ss.
 OS Synthetic.
 XX
 FN WO20020589-A2.
 PD 13-APR-2000.
 PF 30-SEP-1999; 99WO-US022985.
 PR 30-SEP-1998; 98US-0102556P.
 PR 02-OCT-1998; 98US-0102910P.
 PR 21-DEC-1998; 98US-0113229P.
 PR 14-APR-1999; 99US-0129518P.
 XX
 PA (UROG-) UROGENESYS INC.
 PA (AFAR/) AFAR D E.
 PA (HUBE/) HUBERT R S.
 PA (RAIT/) RAITANO A B.
 PA (MITC/) MITCHELL S C.
 PI Afar DE, Hubert RS, Raitano AB, Mitchell SC;
 DR WPI; 2000-317715/27.
 XX
 PT PTAN protein, and sequences encoding them, used for diagnosing and
 PT treating cancers, especially breast and prostate cancers.
 XX
 PS Example 1; Page 31; 71pp; English.
 CC This sequence represents a primer used in the isolation of cDNA fragments
 CC of the PTAN (testis specific protein expressed in prostate cancer) gene.
 CC PTAN is expressed in 3 isoforms PTAN-1, 2, and 3. The PTAN gene is
 CC located on chromosome 1q22. PTAN is overexpressed in prostate cancer, and
 CC has a testis specific expression pattern in adult tissues. PTAN shows no
 CC homology to any known gene. PTAN can be used in methods for the diagnosis
 CC of cancer, especially prostate or breast cancer, where the normal tissue
 CC samples are prostate tissue, or breast tissue, bone tissue, lymphatic
 CC tissue, serum, blood, or urine. A vector containing the PTAN nucleotide
 CC sequence, a vaccine composition targeting PTAN, PTAN, ribozymes specific
 CC for PTAN mRNA and antisense sequences, can be used to treat cancer,
 CC especially breast and prostate cancers. Cancer development can be
 CC inhibited by a vaccine composition targeting PTAN
 XX
 SQ Sequence 42 BP; 3 A; 2 C; 2 G; 35 T; 0 U; 0 Other;
 Query Match 75.2%; Score 18.8; DB 3; Length 42;
 Best Local Similarity 90.9%; Pred. No. 8.6e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTGATC 22
| | | | | | | | | | | | | | | | | | | | | |
Db 25 AAAAAAAAAAAAGCTTGATC 4

Search completed: December 14, 2005, 02:43:19
Job time : 207.2 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 01:34:03 ; Search time 1752.1 Seconds
(without alignments)
667.586 Million cell updates/sec

Title: US-10-681-773-10

Perfect score: 25

Sequence: 1 aaaaaaaaaaagctgacatc 25

Scoring table: IDENTITY NUC

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :
EST:
1: gb_esc1:
2: gb_esc2:
3: gb_esc3:
4: gb_esc4:
5: gb_esc5:
6: gb_esc6:
7: gb_esc7:
8: gb_esc8:
9: gb_esc9:
10: gb_esc10:
11: gb_esc11:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22.4	89.6	321	10	CM797493
2	22.4	89.6	394	1	AV801140
3	22.4	89.6	402	3	BP603671
4	22.4	89.6	419	3	BP608064
5	22.4	89.6	429	1	AV795931
6	22.4	89.6	447	3	BP610473
7	21.8	87.2	709	10	CM631677
8	20.8	83.2	202	7	CNS24164
9	20.8	83.2	282	9	AZ910998
10	20.8	83.2	425	10	CM700831
11	20.8	83.2	478	10	CZ623875
12	20.8	83.2	652	10	AG920908
13	20.2	80.8	180	9	AZ105321
14	20.2	80.8	244	2	BB375648
15	20.2	80.8	250	8	CX065566
16	20.2	80.8	265	7	CO296128
17	20.2	80.8	317	3	BI543181
18	20.2	80.8	356	7	CN072053
19	20.2	80.8	372	6	CF089658
20	20.2	80.8	396	9	AO141946
21	20.2	80.8	449	8	DN799152
22	20.2	80.8	450	3	BM880393

23	20.2	80.8	464	6	CD069749
24	20.2	80.8	473	6	CD069769
25	20.2	80.8	479	9	BH522865
26	20.2	80.8	490	6	CD069753
27	20.2	80.8	491	6	CA563039
28	20.2	80.8	511	6	AO137143
29	20.2	80.8	513	2	BF611985
30	20.2	80.8	517	1	AM618772
31	20.2	80.8	518	5	BQ60514
32	20.2	80.8	524	8	DR102515
33	20.2	80.8	528	9	CC037473
34	20.2	80.8	539	9	AO994802
35	20.2	80.8	549	9	AQ976683
36	20.2	80.8	556	3	BI500321
37	20.2	80.8	558	3	BI500177
38	20.2	80.8	590	9	AZ450294
39	20.2	80.8	618	10	CL735418
40	20.2	80.8	644	9	CE071494
41	20.2	80.8	646	9	AZ450483
42	20.2	80.8	652	9	CE031150
43	20.2	80.8	659	10	CE010279
44	20.2	80.8	663	10	CL718510
45	20.2	80.8	702	8	DR729413
46	20.2	80.8	703	9	AZ989040
47	20.2	80.8	705	10	AG349053
48	20.2	80.8	718	11	CR131244
49	20.2	80.8	757	10	AG571561
50	20.2	80.8	762	8	DR688739
51	20.2	80.8	779	7	CO385790
52	20.2	80.8	787	7	CO084767
53	20.2	80.8	802	10	CL799366
54	20.2	80.8	820	10	CL067511
55	20.2	80.8	853	9	BZ967717
56	20.2	80.8	866	10	CZ225144
57	20.2	80.8	866	11	BCA54323
58	20.2	80.8	876	11	CNS05NCB
59	20.2	80.8	880	9	CC865385
60	20.2	80.8	1088	11	CNS04Y4Q
61	20.2	80.8	1148	9	CC188543
62	20.2	80.8	1249	9	CC324094
63	20.2	80.0	738	10	CL651769
64	20.2	80.0	762	6	CA356715
65	19.8	79.2	204	8	D61775
66	19.8	79.2	251	5	BX614017
67	19.8	79.2	271	11	CR499737
68	19.8	79.2	409	7	CN292167
69	19.8	79.2	479	6	CD451095
70	19.8	79.2	509	2	BE157233
71	19.8	79.2	531	8	CA431362
72	19.8	79.2	542	10	BX238538
73	19.8	79.2	570	6	CD450350
74	19.8	79.2	598	11	DE083880
75	19.8	79.2	604	7	CR535112
76	19.8	79.2	614	5	BX953309
77	19.8	79.2	615	8	CA184687
78	19.8	79.2	632	5	BU618293
79	19.8	79.2	632	6	CA429434
80	19.8	79.2	634	10	CL358313
81	19.8	79.2	664	10	AG055646
82	19.8	79.2	666	9	CE201628
83	19.8	79.2	669	9	BH956495
84	19.8	79.2	671	7	CR629796
85	19.8	79.2	674	10	AG047569
86	19.8	79.2	675	10	AG051677
87	19.8	79.2	683	1	AL599901
88	19.8	79.2	712	3	BM633829
89	19.8	79.2	728	7	CR769513
90	19.8	79.2	779	7	CO393097
91	19.8	79.2	813	7	CA404334
92	19.8	79.2	843	8	CC805815
93	19.8	79.2	848	10	CG943200
94	19.8	79.2	870	5	BU175945
95	19.8	79.2	890	9	BZ166770

CD069749	MA2-0012T
CD069769	MA2-0012T
BH522865	BOHEV13TF
CD069753	MA2-0012T
CA563039	KO308C12-
AO137143	HS 3064_B
BF611985	deB1C01_Y
AM618772	EST320758
BQ60514	H4064G08-
DR102515	STR1_81
CC037473	3591_I_87
AO994802	RPCI-23-3
AQ976683	RPCI-23-3
BI500321	r878e10_Y
BI500177	r878e10_Y
AZ450294	IM0249A09
CL735418	OR_BBA006
CE071494	tigr-g88-
AZ450483	IM0249C15
CE031150	tigr-g88-
CE010279	tigr-g88-
CL718510	OR_BBA004
DR729413	AGENCOURT
AZ989040	2M0272B16
AG349053	Mus muscu
CR131244	Forward s
AG571561	Mus muscu
DR688739	EST107882
CO385790	AGENCOURT
CO084767	GR_EA011
CL799366	OR_CBA001
CL067511	CH216-111
BZ967717	PUBK032TD
CZ225144	A1AA-aaf2
BCA54323	Equus cab
AJ345188	Tetradon
CC865385	NDL_2P12.
AL312515	Tetradon
CC188543	CH261-9K1
CC340494	TAM32-22G
CL651769	PR10113A
CA356715	628958 NC
D61775	HMD209D1B
BX614017	BX614017
CR499737	mtH2-176N
CN292167	170005830
CD451095	USDA-FP_1
BE157233	RC4-HT037
CA431362	JGI_XZ62
BX238538	Danio rer
CD450350	USDA-FP_1
DE083880	Oryzias l
CR535112	CR535112
BX953309	DKFZ781M
CA184687	A11_45-17
BU618293	UI-H-FH1-
CA429434	UI-H-FH1-
CL358313	RPCI44_41
AG055646	Pan trogl
CE201628	tigr-g88-
BH956495	deB1C01_Y
CR629796	DKFZ469G
AG047569	Pan trogl
AG051677	Pan trogl
AL599901	DKFZ313F
BM633829	170006875
CA404334	AUF_1FHK
CR769513	DKFZ469H
CO393097	AGENCOURT
CA404334	AUF_1FHK
CC805815	JGI_CAAJ1
CG943200	BMENNS1TF
BU175945	AGENCOURT
BZ166770	CH230-377

96	19.8	79.2	920	10	CL119775	CL119775 ISB1-7702
97	19.4	77.6	436	3	BP676399	BP676399 BP676399
98	19.4	77.6	448	3	CC945265	CC945265 BOIRB38TF
99	19.4	77.6	472	9	AQ239678	AQ239678 CIT-HSP-2
100	19.4	77.6	531	9	CC952680	CC952680 BOIFN02TF
101	19.4	77.6	567	9	BH510831	BH510831 BOGNB28TF
102	19.4	77.6	573	9	BZ469233	BZ469233 BONFJ26TF
103	19.4	77.6	580	9	BH736678	BH736678 BOMAI17TR
104	19.4	77.6	585	9	BH547486	BH547486 BOGSV32TF
105	19.4	77.6	590	9	CC965165	CC965165 BOIRW86TF
106	19.4	77.6	591	9	BH564768	BH564768 BOHOK83TR
107	19.4	77.6	606	9	BZ460409	BZ460409 BONAP09TR
108	19.4	77.6	632	3	BH567492	BH567492 BOGUG49TF
109	19.4	77.6	647	3	BJ827515	BJ827515 BJ827515
110	19.4	77.6	685	9	BH990584	BH990584 oef82905.
111	19.4	77.6	693	9	BH488293	BH488293 BOGGB68TR
112	19.4	77.6	711	9	BZ006823	BZ006823 oei02h04.
113	19.4	77.6	715	9	BH993161	BH993161 oei61b10.
114	19.4	77.6	735	9	BH602405	BH602405 BOGIB42TF
115	19.4	77.6	737	9	BZ025719	BZ025719 oeb92a02F
116	19.4	77.6	741	8	DR822978	DR822978 EST11451
117	19.4	77.6	749	9	BZ032179	BZ032179 oeh33b11.
118	19.4	77.6	750	9	BZ060581	BZ060581 11907f05.
119	19.4	77.6	752	9	BZ060582	BZ060582 11907f05.
120	19.4	77.6	760	9	BH687058	BH687058 BOHUG79TR
121	19.4	77.6	761	9	BH598221	BH598221 BOGYH92TF
122	19.4	77.6	764	9	BZ017481	BZ017481 oed93e11.
123	19.4	77.6	771	10	CL905016	CL905016 OA_ABA000
124	19.4	77.6	819	10	CL671117	CL671117 PR10163D
125	19.4	77.6	876	8	DR951334	DR951334 EST114287
126	19.4	77.6	1338	10	AJ856529	AJ856529 Brassaica
127	19.4	77.6	1443	10	AJ856528	AJ856528 Brassaica
128	19.2	76.8	160	2	BG603017	BG603017 EST02107
129	19.2	76.8	168	2	BR871173	BR871173 MR1-ET014
130	19.2	76.8	178	6	CB008185	CB008185 VVC054G04
131	19.2	76.8	204	3	BP694695	BP694695 BP694695
132	19.2	76.8	206	9	AZ510748	AZ510748 1M0555L24
133	19.2	76.8	219	7	CK570838	CK570838 est. 1 van
134	19.2	76.8	241	9	CE065540	CE065540 t1gr-g8s-
135	19.2	76.8	264	6	CP421185	CP421185 USDA-FP-1
136	19.2	76.8	274	6	CF262633	CF262633 AUA-1pova
137	19.2	76.8	287	10	CE469394	CE469394 t1gr-g8s-
138	19.2	76.8	295	5	CB490116	CB490116 omvKrtcl10
139	19.2	76.8	300	6	CB492043	CB492043 omvKrtcl10
140	19.2	76.8	319	9	CC555943	CC555943 CC240.463
141	19.2	76.8	344	3	B1326710	B1326710 fps5b01.x
142	19.2	76.8	346	9	AQ106609	AQ106609 HS_3080_B
143	19.2	76.8	364	9	AQ035701	AQ035701 CIT-HSP-2
144	19.2	76.8	372	10	CE537486	CE537486 t1gr-g8s-
145	19.2	76.8	398	1	AL643127	AL643127 AL643127
146	19.2	76.8	398	1	AL785598	AL785598 AL785598
147	19.2	76.8	406	10	CL453302	CL453302 ZMMBBD048
148	19.2	76.8	411	2	BB689885	BB689885 BB689885
149	19.2	76.8	418	9	BH477603	BH477603 BOHCN84TF
150	19.2	76.8	423	9	BZ120103	BZ120103 CH230-472

ALIGNMENTS

RESULT 1
LOCUS CM797493 321 bp DNA linear GSS 23-NOV-2004
DEFINITION WiscDeloxa13-416P2 Arabidopsis thaliana T-DNA insertion flanking
sequences Arabidopsis thaliana genomic, genomic survey sequence.
ACCESSION CM797493
VERSION CM797493.1 GI:5595321
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 321)
AUTHORS Woody,S.T., Monson,S.S., Jester,P.J., Austin-Phillips,S.,
Amadio,R.M., Susman,M.R. and Krysan,P.J.
TITLE A New Community Resource for Knock-Out Small Genes and
Tandemly-Duplicated Gene Families and for Mosaic Analysis in
Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Woody ST
Biotechnology Center
University of Wisconsin-Madison
425 Henry Mall, Madison, WI 53706, USA
Tel: (608) 262-4640
Email: swoody@facstaff.wisc.edu
Class: TAIL-PCR.
FEATURES
source Location/Qualifiers
1..321
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultiivar="Col-O ecotype"
/db_xref="taxon:3702"
/tissue_type="seeds produced by primary (Basta-resistant)
transformation"
/clone_lib="Arabidopsis thaliana T-DNA insertion flanking
sequences"
/note="Vector: pDS-Lox; Sequence generated in the course
of an Arabidopsis T-DNA tagging program. TAIL-PCR was used
to generate sequencing templates that represent A.T.
genomic DNA flanking the left border of the pDS-Lox T-DNA
insert. PCR products were sequenced directly by using the
p745 primer 5' AACGTCGCAAGTGTATTAGTTGTC 3'"

ORIGIN

Query Match 89.6%; Score 22.4; DB 10; Length 321;
Best Local Similarity 95.8%; Pred. No. 4.7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Gy 2 AAAAAAAAAAAGCTGATCTTC 25
Db 238 AAAAAAAAACTAAAGCTGATCTTC 261

RESULT 2
LOCUS AV801140 394 bp mRNA linear EST 29-MAR-2002
DEFINITION AV801140 RAF19 Arabidopsis thaliana cDNA clone RAF109-27-A16 3',
mRNA sequence.
ACCESSION AV801140
VERSION AV801140.1 GI:19835125
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 394)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arai,Kawa,T., Shibata,K., Shingawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinzaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: maseki@rc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda PUC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web

site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

source

```
1..394
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-27-A16"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL9"
/notes="Site_1: BamHI, Site_2: SalI, subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
hr) treatments"
```

ORIGIN

Query Match 89.6%; Score 22.4; DB 1; Length 394;
Best Local Similarity 95.8%; Pred. No. 4.7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
2 AAAAAAAAACTAAGCTTGATCTTC 25

DB 2 AAAAAAAAACTAAGCTTGATCTTC 25

RESULT 3 BP603671 402 bp mRNA linear EST 25-JUN-2004
LOCUS BP603671

DEFINITION BP603671 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-59-J21 3',
mRNA sequence.

ACCESSION BP603671 GI:49233915

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M.,
Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)

JOURNAL

PUBMED

COMMENT

Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
reversed clone; please visit our web site
(<http://pfweb.gsc.riken.go.jp/>) for further details.
Location/Qualifiers

FEATURES

source

```
1..402
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL16-59-J21"
/lab_host="DH10B"
/clone_lib="RAFL16"
/notes="Site_1: BamHI, Site_2: SalI, dark-grown"
```

ORIGIN

Query Match 89.6%; Score 22.4; DB 3; Length 402;
Best Local Similarity 95.8%; Pred. No. 4.7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
2 AAAAAAAAACTAAGCTTGATCTTC 25

DB 23 AAAAAAAAACTAAGCTTGATCTTC 46

RESULT 4

BP608064

LOCUS BP608064 419 bp mRNA linear EST 25-JUN-2004
DEFINITION BP608064 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-77-D23 3',
mRNA sequence.

ACCESSION

BP608064

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T.,
Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M.,
Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A. and Shinozaki, K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)

JOURNAL

PUBMED

COMMENT

Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
reversed clone; please visit our web site
(<http://pfweb.gsc.riken.go.jp/>) for further details.
Location/Qualifiers

FEATURES

source

```
1..419
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL16-77-D23"
/lab_host="DH10B"
/clone_lib="RAFL16"
/notes="Site_1: BamHI, Site_2: SalI, dark-grown"
```

ORIGIN

Query Match 89.6%; Score 22.4; DB 3; Length 419;
Best Local Similarity 95.8%; Pred. No. 4.7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
23 AAAAAAAAACTAAGCTTGATCTTC 46

DB 23 AAAAAAAAACTAAGCTTGATCTTC 46

RESULT 5 AV795931 429 bp mRNA linear EST 29-MAR-2002
LOCUS AV795931/c

DEFINITION AV795931 RAFL8 Arabidopsis thaliana cDNA clone RAFL08-19-N06 3',
mRNA sequence.

ACCESSION

AV795931

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.

Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)

```

COMMENT
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadaei, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekic@tc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda PhiC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
details.

FEATURES
Source
Location/Qualifiers
1..429
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL08-19-N06"
/seq_stage="rossette plants"
/lab_host="DH10B"
/clone_id="RAFL8"
/notice="Site 1: BamHI; Site 2: SalI; subjected to
dehydration-treated (1, 2, 5, 10, 24 hr)"

ORIGIN

Query Match 89.6%; Score 22.4; DB 1; Length 429;
Best Local Similarity 95.8%; Pred. No. 4,7e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 2 AAAAAAAAACTAAAGCTTGATCTTC 25
|||||
Db 152 AAAAAAGAACTAAAGCTTGATCTTC 129
|||||

RESULT 6
BP610473 447 bp mRNA EST 26-JUN-2004
DEFINITION BP610473 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-11-120 3',
mRNA sequence.
ACCESSION BP610473
VERSION BP610473.1 GI:49261655
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; euroside II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 447)
Seki,M., Natsusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
NakaJima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
Hayashizaki,Y., Kawai,Y., Carninci,P., Itoh,M., Ishii,Y.,
Arikawa,T., Shibata,K., Shinozawa,A. and Shinozaki,K.
Functional annotation of a full-length Arabidopsis cDNA collection
Science 296 (5565), 141-145 (2002)
11910074
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadaei, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekic@tc.riken.go.jp
reversed clone; please visit our web site
(http://pfweb.gsc.riken.go.jp/) for further details.
Location/Qualifiers
1..447
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL16-11-120"
/lab_host="DH10B"

FEATURES
Source

```

ORIGIN	Query Match	Best Local Similarity	Matches	Score	DB 3	Length	447
Query Match	89.6%	95.8%	0	0	0	0	0
Best Local Similarity	95.8%	92.0%	0	0	0	0	0
Matches	23	Conservative	0	Mismatches	1	Indels	0
ORIGIN	2	AAAAAAAACTAAGCTTGATCTTC	25				
DB	23	AAAAAAAACTAAGCTTGATCTTC	46				
RESULT 7							
LOCUS	CM631677						
DEFINITION	OP_Ba0058F09.f OP_Ba Oryza punctata genomic clone OP_Ba0058F09						
ACCESSION	CM631677						
VERSION	CM631677.1						
KEYWORDS	GSS.						
SOURCE	Oryza punctata						
ORGANISM	Oryza punctata						
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.						
AUTHORS	Samtuel, P., Westerman, R., Kim, H., Yu, Y., Misosaki, M., Yeast, D., Stum, D., Rao, K., Luo, M., Jett, R., Kudrna, D., Muller, C., Hatfield, J., Soderlund, C., Wang, R., and Jackson, S.A.						
TITLE	OMP Project - Purdue University						
JOURNAL	Unpublished (2004)						
COMMENT	Contact: Scott A. Jackson Jackson Laboratory Purdue University 915 W. State St., West Lafayette, IN 47907, USA Tel.: 765/4963621 Fax: 765/4967255 Email: sjackson@purdue.edu Basecalling by phred version 0.020425.c. This sequence was derived from the raw sequence read by clipping with Lucy version 1.19s. Bases 42-750 of the raw sequence (length 1079) were retained after clipping. PCR Primers FORWARD: TAA TAC GAC TCA CTA TAG GG BACKWARD: CAC TCA TTA GGC ACC CCA Insert length: 161 Std Error: 0.00 Plate: 0058 row: F column: 09 Seq primer: TAA TAC GAC TCA CTA TAG GG Class: BAC ends.						
FEATURES							
source	Location/Qualifiers						
	1..709						
	/organism="Oryza punctata"						
	/mol_type="genomic DNA"						
	/db_xref="taxon:4537"						
	/clone="OP_Ba0058F09"						
	/cmap_type="Young leaves"						
	/lab_host="DH10B-T1 phage resistant"						
	/clone_lib="OP_Ba"						
	/note="Vector: pGIRAC1, Site_1: HindIII, Site_2: HindIII"						
ORIGIN							
Query Match	87.2%	Score 21.8	DB 10	Length 709			
Best Local Similarity	92.0%	Pred. No. 8.1e+02					
Matches	23	Conservative	0	Mismatches	2	Indels	0
ORIGIN	1	AAAAAAAACTAAGCTTGATCTTC	25				
DB	389	AAAAAAAACTAAGCTTGATCTTC	413				

LOCUS CN524164 202 bp mRNA linear EST 28-APR-2004
 DEFINITION GQ015M11.T3_A06 GQ015 *Populus trichocarpa* x *Populus deltoides* cDNA
 accession GQ015M11_A06 5', mRNA sequence.
 VERSION CN524164
 KEYWORDS CN524164.1 GI:46842533
 SOURCE EST.
 ORGANISM *Populus trichocarpa* x *Populus deltoides*
Populus trichocarpa x *Populus deltoides*
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eustosids I; Malpighiales; Salicaceae; Salicaceae; Populus.
 1 (bases 1 to 202)
 Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
 Retel,E. and Mackay,J.
 Arborea EST sequencing in *Populus* sp. (poplar)
 Unpublished (2004)
 JOURNAL Contact: John Mackay
 COMMENT Centre de Recherche en Biologie Forestiere
 Universite Laval
 Pavillon Charles-Bugene Marchand, Quebec, CANADA G1K 7P4
 Fax: 418 656 7493
 Email: jmackay@rsvs.ulaval.ca
 Center for Computational Genomics and Bioinformatics (CCGB),
 University of Minnesota, MN id Identifier: MN5231368 Clone ID:
 GQ015M11_A06 Clones available through: John Mackay, Ph. D.
 Professeur adjoint - Assistant professeur EMAIL:
 jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
 (Forest Biology Research Center) Universite Laval Quebec, Quebec
 CANADA G1K 7P4
 Plate: M11 row: 06 column: A
 Seq primer: T3 primer.
 FEATURES
 source Location/Qualifiers
 1..202
 /organism="Populus trichocarpa x Populus deltoides"
 /mol_type="mRNA"
 /strain="H11-11"
 /db_xref="taxon:3695"
 /clone="GQ015M11_A06"
 /sex="Male"
 /tissue_type="Shoot tips and secondary stems of trees
 approximately 80 cm tall"
 /lab_host="E. coli DH10B cells"
 /clone_11b="GQ015"
 /note="Organ: Shoot tips and secondary stems of trees
 approximately 80 cm tall; Vector: pBluescript II SK (+)
 XR; Site 1: Eco-RI; Site 2: Xho-I; Shoot tips (including
 apex, primary stem and developing leaves up to and
 including LPI 1) and whole stems (wood and bark)
 undergoing secondary growth taken from between LPI 8 and
 LPI 15. Pooled sample of tissues harvested every two hours
 during one 12 hour light/12 hour dark cycle, beginning 1
 hour before the beginning of the light period and ending 3
 hours after the end of the light cycle. cDNA was prepared
 from 5 Ncg of poly A+ selected RNA and was directionally
 ligated into the pBluescript II SK (+) XR vector
 (Stratagene), transformed by electroporation into DH10B
 cells (in vitro) for propagation"

ORIGIN
 Query Match 83.2%; Score 20.8; DB 7; Length 202;
 Best Local Similarity 91.7%; Pred. No. 1.8e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAAGCTTGATCTT 24
 |||||
 Db 198 AAAAAAAAAAGCAAGCTTGATCTT 175
 |||||
 RESULT 9 282 bp DNA linear GSS 05-MAR-2001
 A2910928
 LOCUS A2910928/c
 DEFINITION RPCI-24-114F16.TVB RPCI-24 Mus musculus genomic clone
 RPCI-24-114F16, genomic survey sequence.

ACCESSION A2910928
 VERSION A2910928.1 GI:13229873
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 282)
 Zhao,S., Niemman,W., Malek,J., Shatsman,S., Akintet,B., Levins,M.,
 Tsengaye,G., Geer,K., Krol,M., Shavrebeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 JOURNAL Other GSSs: RPCI-24-114F16.TVB
 COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@igr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end
 page: http://www.choi.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 114 row: F column: 16
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..282
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-114F16"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /clone_11b="RPCI-24"
 /note="Vector: pTRABAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTRABAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

ORIGIN
 Query Match 83.2%; Score 20.8; DB 9; Length 282;
 Best Local Similarity 91.7%; Pred. No. 1.9e+03;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAAAAAAAACTAAGCTTGATCTT 24
 |||||
 Db 40 AAAAAAAAAAGCAAGCTTGATCTT 17
 |||||
 RESULT 10 425 bp DNA linear GSS 04-NOV-2004
 CW700431
 LOCUS CW700431/c
 DEFINITION A1AA-aab3g03.b1 *Ancylostoma caninum* whole genome shotgun library
 (A1AAGSS 001) *Ancylostoma caninum* genomic, genomic survey sequence.
 accession CW700431
 VERSION CW700431.1 GI:55383007
 KEYWORDS GSS.
 SOURCE *Ancylostoma caninum* (dog hookworm)
 ORGANISM *Ancylostoma caninum*
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
Ancylostomatoidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 1 (bases 1 to 425)
 Mltreva,M., McGarther,J.P., Pape,D., Rittner,E., Tsagarishvili,R.,
 Ronko,I., Martin,J., Wyllie,T., Dante,M., Meyer,R., Messina,D.,
 Waterston,R.H., Clifton,S.W. and Wilson,R.
 Genome Survey sequences from the parasitic nematode *Ancylostoma*

```
1. .478
/organism="Oryza minuta"
/mol_type="genomic DNA"
/db_xref="taxon:63629"
```

Query Match	83.2%	Score 20.8;	DB 10;	Length 652;
Best Local Similarity	91.7%;	Pred. No. 1.9e+03;		
Matches 22;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0

Oy	1	AAAAAAAAAAGCTGATCTT	24
Dd	428	AAAAAAAAAAGCTGATCTT	451
RESULT 13			
LOCUS	AZ105321	180 bp	DNA linear GSS 09-MAY-2000
DEFINITION	RPCI-23-22A8.TJ RPCI-23 Mus musculus genomic clone RPCI-23-22A8,		
ACCESSION	AZ105321		
VERSION	AZ105321.1 GI:7758376		
KEYWORDS	GSS.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 180) Zhao,S., Niemann,W., Feldblum,T., Malek,J., Shatman,S., Almstet,B., Levins,M., McGinn,S., Tesgaye,G., Geer,K., Krol,M., de Jong,P. and Frazer,C.M. Mouse BAC End Sequences from Library RPCI-23 Unpublished (1999) Other GSSs: RPCI-23-22A8.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html Plate: 22 row: A column: 8 Seq primer: SP6 Class: BAC ends.		
FEATURES			
source	Location/Qualifiers		
	1..180		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="RPCI-23-22A8"		
	/sex="Female"		
	/lab_host="DH10B"		
	/clone_lib="RPCI-23"		
	/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Site selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)." "		
ORIGIN			
Query Match	80.8%; Score 20.2; DB %;	Length 180;	
Best Local Similarity	88.0%; Pred.No.3.le+03;		
Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
Oy	1	AAAAAAAAAAGCTGATCTTC	25
Dd	11	AAAAACAACAAATCTGATCTTC	35
RESULT 14			
LOCUS	BH375648/c	244 bp	mRNA linear EST 13-JUL-2000

DEFINITION	BB375648 Riken full-length enriched, 16 days embryo head Mus musculus cDNA clone C130081H05.3, similar to M96853 Rat postnaptic density protein (PSD-95), homologue of discs-large tumor suppressor protein mRNA, mRNA sequence.
ACCESSION	BB375648
VERSION	BB375648.1
KEYWORDS	GI:9088462
SOURCE	EST
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus. 1 (bases 1 to 244)
AUTHORS	Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Iizawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saico,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toyota,T., Tsunoda,Y., Watabiki,A., Matsunabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshinide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.jp, URL:http://genome.gsc.riken.jp/ Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoke,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Thermosensitization and thermotaxis of the thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh,M., Kikuchi,N., Akiyama,J., Shibata,K., Iizawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.riken.go.jp) for further details. Location/Qualifiers 1..244 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="C130081H05" /sex="mixed" /tissue_type="head" /dev_stage="16 days embryo" /lab_host="DH10B" /clone_1fb="RIKEN full-length enriched, 16 days embryo head" /note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'-GAGGAGAGAGAGATCCAAAGGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using triphase thermo-activated reverse

transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTATTAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I."

ORIGIN

Query Match 80.8%; Score 20.2; DB 2; Length 244;
Best Local Similarity 88.0%; Pred. No. 3.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||
31 AAAAAAAAACTAAGCTTGACCTTC 7

RESULT 15
CX066556 250 bp mRNA linear EST 03-JAN-2005
LOCUS 1321175 NCCCA 04RT Oncorhynchus mykiss cDNA, mRNA sequence.
ACCESSION CX066556
VERSION CX066556.1 GI:56988122
KEYWORDS EST.
SOURCE Oncorhynchus mykiss (rainbow trout)
ORGANISM Oncorhynchus mykiss

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 250)
Rexroad, C.E., Rise, M., Koop, B., von Schalburg, K. and Yao, J.

04RT gpool, NCCCA/WVU EST Project, Phase II, in collaboration with
GRASP

Unpublished (2004)
Contact: Rexroad CE
USDA, ARS, National Center for Cool and Cold Water Aquaculture
11876 Leetown Road, Kearneysville, WV 25430, USA
Tel: 304 724 8340 x2129
Fax: 304 725 0351

Email: cirexroad@nccca.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.

Plate: 104 row: L column: 21
Seq primer: ATTAGGTGACCTATAG.

FEATURES

Source

1..250
Location/Qualifiers
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/lab_host="TOP10"
/clone_lib="NCCCA 04RT"
/note="Vector: PCR 4-TOPO; This is an early neurogenesis
SSH library created by Mathew L. Rise constructed by
subtracting late neurogenesis (mixed stages: hindbrain
swelling + heart tube with peristalsis) from early
neurogenesis (mixed stages: neural groove + 1/2 epiboly).
Fish were from a domesticated strain (Spring Valley Trout
Farm, Langley, B.C.), courtesy of Bob Devlin, DFO. These
are mostly internal (coding) sequences."

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 250;
Best Local Similarity 88.0%; Pred. No. 3.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||
219 AAAAAAAAAATAAGCTTGACCTGC 243

RESULT 16

CO296128/c

LOCUS CO296128 265 bp mRNA linear EST 25-JUN-2004
DEFINITION EK212763.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
melanogaster cDNA clone EK212763 5, mRNA sequence.

ACCESSION CO296128

VERSION CO296128.1 GI:49217759

KEYWORDS EST.

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 265)

Kopczynski, C., Platt, D., Campbell, J., Muzong, C., Laufer, A.,
Peterson, E. and Swimmer, C.

Exelixis FlyTag EST Project CK01 Library

Unpublished (2004)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
Plate: EK.2127 row: F column: 3

High quality sequence stop: 264.

FEATURES

Source

1..265
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone_lib="EK212763"
/clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
/note="Organism: mixed stage embryos, imaginal disks, and
adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
XhoI; Random primed, normalized library from mixed stage
embryos, imaginal disks, and adult heads."

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 265;
Best Local Similarity 88.0%; Pred. No. 3.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||
121 AAAAAAAAACTAAGCTTGATATTC 97

RESULT 17

LOCUS B1543181

DEFINITION B1543181 317 bp mRNA linear EST 04-SEP-2001
949075C10.X1 949 - Juvenile leaf and shoot cDNA from Steve Moose
Zea mays cDNA, mRNA sequence.

ACCESSION B1543181

VERSION B1543181.1 GI:15427359

KEYWORDS EST.

SOURCE

ORGANISM

Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 317)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949075 row: C column: 10.

FEATURES

SOURCE

Location/Qualifiers
1..317

/organism="Zea mays"
/mol_type="mRNA"

/cultivar="W64A"

/db_xref="taxon:4577"

/tissue_type="immature leaf primordium and vegetative

meristem

/dev_stage="4 stages from 3-13 days after imbibing"

/lab_host="E. coli XL0LR"

/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve

Moore"

/note="Organ: juvenile vegetative shoots; Vector:

pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts

of total RNA by weight from 4 tissue sources (see below)

were pooled, polyA+ RNA isolated, and cDNA synthesized for

EcoRI (5') and XhoI (3') directional cloning into lambda

Hybridzap vector from Stratagene. Tissue Sources: 1. Whole

shoots 3 days after sowing/imbibing in wet soil. 2. Basal

1.5 cm shoots 6 days after sowing - includes yellow

portions of developing leaves 1-5, primordia from 6-8, and

the vegetative apex. 3. Non-green portions of developing

leaves 4-5 and the vegetative apex, including adult leaf

primordia, 9 days after sowing. 4. Partially expanded and

greening leaves 4-5 at 13 days after sowing."

ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 3; Length 317;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGATCTTC 25

Db 114 AAAAAAAAAGTAATCTTGATCTCC 138

RESULT 18

LOCUS

CN072053 356 bp mRNA linear EST 30-MAR-2004

DEFINITION 1021024D04.x1 1021 - Unigene II from Maize Genome Project Zea mays

CDNA, mRNA sequence.

ACCESSION CN072053

VERSION CN072053.1 GI:45848110

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Place: 1021024 row: D column: 04.

Location/Qualifiers

1..356

/organism="Zea mays"

/mol_type="mRNA"

/db_xref="taxon:949075C10.x1"

/db_xref="taxon:4577"

/clone_lib="1021 - Unigene II from Maize Genome Project"

/note="This library represents the unique genes found in

the second round of EST sequencing at Stanford University

for the maize genome project. Sequences are present from

library 949. Contigs were assembled using ZmBAssembler

ORIGIN

Query Match 80.8%; Score 20.2; DB 7; Length 356;

Best Local Similarity 88.0%; Pred. No. 3.1e+03;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGATCTTC 25

Db 157 AAAAAAAAAGTAATCTTGATCTCC 181

RESULT 19

LOCUS

CF089658 372 bp mRNA linear EST 22-JUL-2003

DEFINITION QHM23J04.yg.ab1_QH_M sunflower H. argophyllus Helianthus argophyllus

CDNA clone QHM23J04, mRNA sequence.

ACCESSION CF089658

VERSION CF089658.1 GI:33128725

KEYWORDS EST.

SOURCE Helianthus argophyllus

ORGANISM Helianthus argophyllus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

asterids; campanulids; Asterales; Asteraceae; Asteroideae;

Heliantheae; Helianthus.

1 (bases 1 to 372)

Kozik, A., Michelmore, R. W., Knapp, S., Matvienko, M., Rieseberg, L.,

Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, U.,

Ellison, P., Kolkmann, J., Slabaugh, M. S., Livingston, K., Zhou, Y.,

Lai, Z., Church, S., Jackson, L. and Bradford, K.

Lettuce and Sunflower ESTs from the Compositae Genome Project

http://compositae.ucdavis.edu/

Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@ucdavis.edu]

belongs to contig QH_Ca_Contig475, see http://cgpdb.ucdavis.edu/

for details.

Plate: QHM23 row: J column: 04.

Location/Qualifiers

1..372

/organism="Helianthus argophyllus"

/mol_type="mRNA"

/db_xref="taxon:73275"

/clone="QH23J04"

/lab_host="E.coli"

/clone_lib="QH_M sunflower H. argophyllus"

/note="Vector: pRCDNA511A; The library was constructed

from three different sources (seedling, root and leaf) of

RNA from a single genotype. cDNAs were pooled and

directionally cloned into a custom medium-copy vector.

Details of library construction can be obtained at

http://cgpdb.ucdavis.edu/

ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 372;

Best Local Similarity 88.0%; Pred. No. 3.1e+03;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAAAAAAAACTTAAGCTTGATCTTC 25

Db 345 AAAAAAAAAGTAATCTTGATCTAC 369

RESULT 20

/db_xref="taxon:6306"
/dev_stage="egg"
/lab_host="DH10B"
/clone_lib="Meioidogyne incognita egg PAMP1 Bird"
/note="Vector: PAMP1 (Gibco) site 1: Not; Site 2: Sal;
The library was provided by Dr. David Bird at North
Carolina State University, Raleigh, NC
(david.bird@ncsu.edu). The cDNA was made by using Dynabead
oligo-dT priming (Dyna1). PCR based library using a
modified protocol from the SMART PCR cDNA Synthesis Kit
from Clontech. Directionally cloned into the UDG sites of
PAMP1."

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 450;
Best Local Similarity 88.0%; Pred. No. 3.1e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
Db 130 AAAAAAAAAATTAATCTTCATCTTC 154

RESULT 23

LOCUS CD069749 464 bp mRNA linear EST 14-SEP-2003
DEFINITION MA2-0012T-V071-E01-U.B MA2-0012 Schistosoma mansoni cDNA clone
ACCESSION CD069749
VERSION CD069749.1 GI:34620842

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatidae; Schistosomidae; Schistosoma.
1 (bases 1 to 464)

REFERENCE

AUTHORS

Verjovski-Almeida, S., Demarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y., Jr.,
Kiteajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Lette, R.A., Malaguas, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni

TITLE

Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni

JOURNAL

PUBMED

COMMENT

Net. Genet. 35 (2), 148-157 (2003)
12973350
Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjovski@iq.usp.br
This sequence was derived from the PABSP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: MA2-0012T-V071 row: 1 column: E.
Location/Qualifiers

FEATURES

source

1..464
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MA2-0012T-V071-E01.B"
/sex="mixed pool"
/dev_stage="adult"
/lab_host="Mus musculus"
/clone_lib="MA2-0012"
/note="Vector: Topoblunt"

ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 464;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
Db 80 AACAAAGAACTTAAGCTGATCTTC 104

RESULT 24

LOCUS CD069769 473 bp mRNA linear EST 14-SEP-2003
DEFINITION MA2-0012T-V071-G01-U.B MA2-0012 Schistosoma mansoni cDNA clone
ACCESSION CD069769
VERSION CD069769.1 GI:34620862

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatidae; Schistosomidae; Schistosoma.
1 (bases 1 to 473)

REFERENCE

AUTHORS

Verjovski-Almeida, S., Demarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y., Jr.,
Kiteajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Lette, R.A., Malaguas, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni

TITLE

Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni

JOURNAL

PUBMED

COMMENT

Net. Genet. 35 (2), 148-157 (2003)
12973350
Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjovski@iq.usp.br
This sequence was derived from the PABSP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: MA2-0012T-V071 row: 1 column: G.
Location/Qualifiers

FEATURES

source

1..473
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MA2-0012T-V071-G01.B"
/sex="mixed pool"
/dev_stage="adult"
/lab_host="Mus musculus"
/clone_lib="MA2-0012"
/note="Vector: Topoblunt"

ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 473;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
|||||
Db 80 AACAAAGAACTTAAGCTGATCTTC 104

RESULT 25

BH522865 LOCUS 479 bp DNA linear GSS 13-DEC-2001
 DEFINITION BOHEV13TR BOHE Brassica oleracea genomic clone BOHEV13, genomic survey sequence.
 ACCSSION BH522865
 VERSION BH522865.1 GI:17730950
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 479)
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S., Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
 Whole genome shotgun sequencing of *Brassica oleracea* and its application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)
 JOURNAL PUBMED 15805490
 COMMENT Other GSSs: BOHEV13TR
 Contact: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: ctown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.
 Location/Qualifiers
 1..479
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO100DH3"
 /db_xref="taxon:3712"
 /clone_lib="BOHEV13"
 /clone_1ib="BOHE"
 /note="Vector: PHOS1; Site 1: BatX1; 2-3 kb sheared genomic DNA inserted into PHOS1 using BatX1 linkers"
 ORIGIN
 Query Match 80.8%; Score 20.2; DB 9; Length 479;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 1 AAAAAAAAACTAAGCTTGATCTTC 25
 191 AAAAAAAAAATTAAAGCTTGATTTTC 215
 RESULT 26
 LOCUS CD069753 490 bp mRNA linear EST 14-SEP-2003
 DEFINITION MA2-0012T-V071-E06-U.B MA2-0012 Schistosoma mansoni cDNA clone
 MA2-0012T-V071-E06.B, mRNA sequence.
 ACCSSION CD069753
 VERSION CD069753.1 GI:34620846
 KEYWORDS EST.
 SOURCE Schistosoma mansoni
 ORGANISM Schistosoma mansoni
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 Strigeida; Schistosomatidae; Schistosomatidae; Schistosoma.
 1 (bases 1 to 490)
 Verjovski-Almeida,S., Demarco,R., Martins,E.A.L., Guimaraes,P.E.M., Ojopi,E.P.B., Paquola,A.C.M., Plazza,J.P., Nishiyama,M.Y., Jr., Kiteajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldi,M.F., Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L., Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A., Nascimento,A.L.T.O., Ohweiler,F.P., Reis,E.M., Ribeiro,M.A., Sa,R.G., Stukerc,G.C., Soares,M.B., Garioni,C., Kawano,T., Rodrigues,V., Medeira,A.M.B.N., Wilson,R.A., Wenck,C.F.M., Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
 Transcriptome analysis of the acelomate human parasite Schistosoma

JOURNAL Mansoni
 Nat. Genet. 35 (2), 148-157 (2003)
 PUBMED 12973350
 COMMENT Contact: Dr. Sergio Verjovski-Almeida
 Departamento de Biocimica
 Instituto de Quimica - Universidade de Sao Paulo
 Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
 Brasil
 Tel: +55-11-3091-2173
 Fax: +55-11-3091-2186
 Email: verjovski@usp.br
 This sequence was derived from the PAPESP Schistosoma mansoni EST
 Genome Project. All sequences in the project were assembled and
 annotated. This entry and all the assembled sequences can be seen
 in the following URL <http://bioinfo.iq.usp.br/schisto/>
 Plate: MA2-0012T-V071 row: 6 column: E.
 Location/Qualifiers
 1..490
 /organism="Schistosoma mansoni"
 /mol_type="mRNA"
 /db_xref="taxon:6183"
 /clone="MA2-0012T-V071-E06.B"
 /sex="mixed pool"
 /dev_stage="adult"
 /lab_host="Mus musculus"
 /clone_lib="MA2-0012"
 /note="Vector: topoblunt"
 ORIGIN
 Query Match 80.8%; Score 20.2; DB 6; Length 490;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 1 AAAAAAAAACTAAGCTTGATCTTC 25
 80 AACAAAGAACTAAGCTTGATCTTC 104
 RESULT 27
 LOCUS CA563039 491 bp mRNA linear EST 19-NOV-2002
 DEFINITION K0308C12-5N NIA Mouse Osteoblast cDNA library (Long) Mus musculus
 cDNA clone NIA:K0308C12 IMAGE:30055043 5', mRNA sequence.
 ACCSSION CA563039
 VERSION CA563039.1 GI:25107694
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 491)
 Piao,Y., Kargul,G.D., Dudekula,D.B., Qian,Y., Luo,A., Carter,M.G., Umezawa,A. and Ko,M.S.H.
 Systematic Analyses of NIA Mouse Osteoblast cDNA Library (Long)
 Unpublished (2001)
 Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cnaa@gsun.grc.nia.nih.gov
 Plate: K0308 row: C column: 12
 Seq primer: M13 Reverse
 High quality sequence stop: 491
 POLYA=NO.
 Location/Qualifiers
 1..491
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C3H/He"
 /db_xref="EBS:K0308C12-5N"
 /db_xref="taxon:10090"
 /clone="NIA:K0308C12 IMAGE:30055043"


```

/issue_type="Osteoblast"
/cell_line="KUSA-A1 cells"
/lab_host="DH10B"
/clone_lib="NIA Mouse Osteoblast cDNA Library (long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI. Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://1gsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Dr. Akihito Umezawa (Keio University School
of Medicine, Japan). Double-stranded cDNAs were
synthesized with an oligo(dT) primer [Invitrogen:
5'-TGACTAGTTCATGATCGAGCGCGCCCTTTTCTTTTCTT-3'] from
2.1 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to lone-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 3.0 kb. The library was constructed
by Yulan Piao (NIA)."
```

ORIGIN

```

Query Match      80.8%; Score 20.2; DB 6; Length 491;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

Oy 1 AAAAAAAAACTAAGCTGATCTTC 25
    ||||| ||||| ||||| |||||
Db 242 AAAAAAAAACAAAAGCGTATTTTC 266
```

```

RESULT 28      A0137143      511 bp      DNA      linear      GSS 24-SEP-1998
LOCUS      A0137143
DEFINITION      HS_3064_B2_E06_MF CIT Approved Human Genomic Sperm Library D Homo
sequence.
ACCESSION      A0137143
VERSION      A0137143
KEYWORDS
SOURCE      GSS.
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo.
1 (bases 1 to 511)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10493764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3064 row: J column: 12
Class: BAC ends
High quality sequence stop: 511.
Location/Qualifiers
1..511
```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT
```

```

FEATURES
source
```

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=3064 Col=12 Row=J"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/notes="Organ: sperm; Vector: pReloBAC11; BAC Clones in
E-Coli DH10B"
```

ORIGIN

```

Query Match      80.8%; Score 20.2; DB 9; Length 511;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

Oy 1 AAAAAAAAACTAAGCTGATCTTC 25
    ||||| ||||| ||||| |||||
Db 247 AAAAAAAAACTAAGCTGATCTTC 271
```

```

RESULT 29      BF611985      513 bp      mRNA      linear      EST 14-DEC-2000
LOCUS      BF611985
DEFINITION      deict01.y1 Wellcome CRC prn3 St19 26 Xenopus laevis cDNA clone
IMAGE:3548136 5', mRNA sequence.
ACCESSION      BF611985
VERSION      BF611985.1 GI:11782169
KEYWORDS
SOURCE      EST.
ORGANISM      Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 513)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L.,
Pape,D., Martin,J., Wyllie,T., Underwood,K., Theising,B., Bowers,Y.,
Peterson,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Maenu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Maenu Xenopus EST project, 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, K. Ryan and A.M. Zorn
(Wellcome/CRC Institute). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/INMIL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 504.
Location/Qualifiers
1..513
```

```

FEATURES
source
```

```

/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:9355"
/clone="IMAGE:3548136"
/tissue_type="pooled embryos, stage 19-26"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Wellcome CRC prn3 St19 26"
/notes="Vector: pBSRN3; Site 1: NotI; Site 2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library was constructed
by N. Garrett, E. Bellefroid, and A.M. Zorn.
(Wellcome/CRC Institute)."
```

ORIGIN

```

Query Match      80.8%; Score 20.2; DB 2; Length 513;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Oy 1 AAAAAAAAACTAAGCTTGCATCTTC 25
 |||||
 Db 447 AAAAAAAAAATAACTTGCATCTTC 471

RESULT 30
 LOCUS AM618772/c
 DEFINITION EST320758 L. pennellii trichome, Cornell University Lycopersicon
 AM618772
 VERSION AM618772.1 GI:7325018
 KEYWORDS EST.
 SOURCE Lycopersicon pennellii (Solanum pennellii)
 ORGANISM Lycopersicon pennellii
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 517)
 Alcala, J., Vrebalov, J., White, R., Matern, A.L., Lakey, J., Holt, I.E., Liang, F., Hansen, T.S., Upson, D., Romling, C.M., Claven, M.B., Fujii, C.Y., Bowman, C.L., Niernan, W., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
 Generation of ESTs from wild tomato (*Lycopersicon pennellii*) trichomes
 Unpublished (1999)
 CONTACT: CUGI
 JOURNAL: Clemson University Genomics Institute
 COMMENT: Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES
 source
 location/Qualifiers
 1..517
 /organism="Lycopersicon pennellii"
 /mol_type="mRNA"
 /db_xref="taxon:28526"
 /clone="CLPT16M11"
 /tissue_type="trichome"
 /dev_stage="mixed stages"
 /lab_host="SOLK"
 /clone_idb="L. pennellii trichome, Cornell University"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; leaves of various stages were shaken in liquid nitrogen, shearing off trichomes. This procedure yielded a mixture of cells highly enriched for trichomes, with minor contamination by other types of leaf cells."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 1; Length 517;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGCATCTTC 25
 |||||
 Db 48 AAAAAAAAACAAACTTGTCTTC 24

RESULT 31
 LOCUS B0560514
 DEFINITION H4064G08-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
 H4064G08 5', mRNA sequence.
 ACCESSION B0560514
 VERSION B0560514.1 GI:21461399
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 518)
 VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G.,

TITLE
 JOURNAL
 COMMENT

RESULT 32
 LOCUS DR102515
 DEFINITION STR1_81 E05_g1 A033 Stem Response Resistant Pinus taeda cDNA clone
 STR1_81 E05_A033 5', mRNA sequence.
 ACCESSION DR102515
 VERSION DR102515.1 GI:67561870
 KEYWORDS EST.
 SOURCE Pinus taeda (loblolly pine)
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 1 (bases 1 to 524)
 Neale, D., Cordominier-Pratt, M.M., Morse, A., Davis, J., Lorenz, W.W., Pratt, L., and Dean, J.F.D.
 An EST database from pitch canker resistant loblolly pine (*Pinus taeda*) shoot tips
 Unpublished (2005)
 JOURNAL: Other ESTs: STR1_81 E05_g1 A033
 CONTACT: Cordominier-Pratt MM
 LABORATORY: Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmprat@uga.edu
 Forest Resources, University of Georgia; plant material prepared by Allison M. Morse (School of Forest Resources and Conservation,

FEATURES
 source
 location/Qualifiers
 1..518
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="H4064G08-5"
 /dev_stage="mixed"
 /sex="mixed"
 /lab_host="DH10B"
 /clone_idb="NIA Mouse 7.4K cDNA Clone Set"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI. This clone is among a rearranged set of 7,407 clones from more than 20 cDNA libraries."

ORIGIN
 Query Match 80.8%; Score 20.2; DB 5; Length 518;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGCATCTTC 25
 |||||
 Db 242 AAAAAAAAACAAACTTGTCTTC 266

University of Florida) using hedges provided by the Forest Biology Research Cooperative (FBRC) and the CCLONES project at the University of Florida; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia.
Seq primer: JENREV (CAGGAACAGCTATGACC).

FEATURES

source

Location/Qualifiers

1..524

/organism="Pinus taeda"

/mol_type="mRNA"

/strain="3 CCLONES"

/db_xref="taxon:3352"

/clone="STR1_81.R05.A033"

/lab_host="DH10B-T1 phage-resistant E. coli"

/clone_lib="Stem Response Resistant"

/note="Organ: Shoot tip; Vector: pSH180; Site 1: EcoRI;

Site 2: XhoI; The library was prepared from polyA+ RNA

isolated from stems of loblolly pine (Pinus taeda) hedges

circumcinctum, the inciting agent of pitch canker disease.

Four pitch canker resistant hedges were inoculated in

November, 2003. For each hedge, multiple succulent shoot

tips were inoculated with 1 mL of a 500 spores/mL solution

after excision. The terminal 1 cm of each challenged stem

was harvested for mRNA isolation 10 days after challenge

ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 8; Length 524;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

1 AAAAAAAAACTAAAGCTGATCTTC 25

Db

295 AAGGAAAACTAAAGCTGATCTTC 319

RESULT 33

CC037473

LOCUS

3591.1 87.1 H03.Y.1 3591 - RescuenMu Grid P Zea mays genomic,
genomic survey sequence.

ACCESSION

CC037473.1 GI:29452364

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

1 (bases 1 to 528)

Maize genomic sequences found using engineered RescuenMu transposon

Unpublished (2001)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 3591_1_87_1 row: 22

Classes: transposon-tagged,
Location/Qualifiers

1..528

/organism="Zea mays"

/mol_type="genomic DNA"

/cultivar="mixed Background W23/A188/B73/K55"

/db_xref="taxon:4577"

/tissue_type="leaf"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="3591 - RescuenMu Grid P"

/note="Organ: leaf; Vector: RescuenMu (engineered from

ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 9; Length 528;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db

1 AAAAAAAAACTAAAGCTGATCTTC 25

65 AAAAAAAAACTAAAGCTGATCTTC 89

RESULT 34

AQ994802

LOCUS

AQ994802 RPCT-23-383116.TV RPCT-23 Mus musculus genomic clone
RPCT-23-383116, genomic survey sequence.

ACCESSION

AQ994802

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other-GSS: RPCT-23-383116.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaod@igf.org

Clones are derived from the mouse BAC library RPCT-23. For BAC

library availability, please contact Pieter de Jong

(pieter@jlong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 383 row: 1 column: 16

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..539

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCT-23-383116"

/sex="female"

/lab_host="DH10B"

/clone_lib="RPCT-23"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:

EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the

pBACScript backbone); Site 1: BamHI, Site 2: BglII;
RescuenMu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on RescuenMu, go to the web
site 'www.zmdb.iastate.edu' and follow the links for
'RescuenMu.' Grid P was grown at Moltokai in 2002. DNA was
extracted from leaf strips, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin."

ECORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 539;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAAAAAAAACTAAGCTTGATCTTC 25
12 AAAAAAAAACAAAATCTTGATCTTC 36

Db

RESULT 35

LOCUS

DEFINITION A0976683 549 bp DNA linear GSS 29-JAN-2000
RPCI-23-334C19.TV RPCI-23 Mus musculus genomic clone
A0976683
A0976683 genomic survey sequence.

ACCESSION A0976683.1 GI:6808984

VERSION

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoe@igf.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACDAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea.ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 334 Row: C Column: 19
Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..549
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-334C19"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1: EcorI; Site 2: EcorI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcorI and EcorI MethyIase. Size selected DNA was cloned into the pBAC3.6 vector at the EcorI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 549;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAAAAAAAACTAAGCTTGATCTTC 25
12 AAAAAAAAACAAAATCTTGATCTTC 36

Db 18 AAAAAAAAACAAAATCTTGATCTTC 42

RESULT 36

LOCUS

DEFINITION B1500321 556 bp mRNA linear EST 28-AUG-2001
r788e10.y1 Prist pacificus mixed stage PAMPI v1 Chlapelli McCarter
Pristionchus pacificus cDNA 5' similar to SM:TBA2_CASEL P34690
TUBULIN ALPHA-2 CHAIN. [1] ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LMLT; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
High quality sequence stop: 1.

FEATURES

source

Location/Qualifiers
1..556
/organism="Pristionchus pacificus"
/mol_type="mRNA"
/db_xref="taxon:54126"
/dev_stage="mixed"
/lab_host="DH10B"
/clone_lib="Prist pacificus mixed stage PAMPI v1 Chlapelli McCarter"
/note="Vector: pAMP1 (Gibco); Site 1: NotI; Site 2: SalI; The library was constructed by Brandt Chlapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dynal). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of PAMPI. "

ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 556;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAAAAAAAACTAAGCTTGATCTTC 25
193 AAAAAAAAACTGAGCTTGCTCTTC 169

Db

RESULT 37

LOCUS

DEFINITION B1500177 558 bp mRNA linear EST 28-AUG-2001
r786f09.y1 Prist pacificus mixed stage PAMPI v1 Chlapelli McCarter
Pristionchus pacificus cDNA 5' similar to SM:TBA2_CASEL P34690
TUBULIN ALPHA-2 CHAIN. [1] ;, mRNA sequence.

ACCESSION B1500177
VERSION B1500177.1 GI:15339521
KEYWORDS EST.

SOURCE
Pristionchus pacificus

ORGANISM
Pristionchus pacificus

REFERENCE
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Diplogasteridae; Pristionchus.
1 (bases 1 to 558)

AUTHORS
McCarter,J., Clifton,S., Chiapelli,B., Page,D., Martin,J.,
Wylie,T., Dantle,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
Boveri,Y., Gibbons,M., Ritzer,E., Bennett,U., Franklin,C.,
Tsagarisvilli,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
The Washington Univ. Nematode EST Project, 1999

TITLE
Unpublished (1999)

JOURNAL
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999

COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
High quality sequence stop: 369.

FEATURES
source
1..558
Location/Qualifiers
/organism="Pristionchus pacificus"
/mol_type="mRNA"
/db_xref="taxon:54126"
/dev_stage="mixed"
/lab_host="DH10B"
/clone_lib="Prist pacificus mixed stage PAMPI v1 Chiapelli1
McCarter"
/note="Vector: PAMPI (Gibco); Site 1: NotI; Site 2: SalI;
The library was constructed by Brandi Chiapelli and Dr.
James McCarter at Washington University, St. Louis. The
cDNA was made by using Dynabead oligo-dT priming (Dynal).
PCR based library using a modified protocol from the
SMART PCR cDNA Synthesis Kit from Clontech. Directionally
cloned into the UDG sites of pAMPI. "

ORIGIN
Query Match 80.8%; Score 20.2; DB 3; Length 558;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 AAAAAAAAACTAAGCTTGATCTTC 25
195 AAAAAAAAACTGAGCTTCTTC 171

RESULT 38
AZ450294 590 bp DNA linear GSS 04-OCT-2000
LOCUS 1M0249A09F Mouse 10kb plasmid UGCLM library Mus musculus genomic
DEFINITION clone UGCLM0249A09 F, genomic survey sequence.
ACCESSION AZ450294
VERSION AZ450294.1 GI:10605127
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 590)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddm@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0249 row: A column: 09
Seq primer: CGTGTAAACGACGCACG
Class: plasmid ends
High quality sequence stop: 590.

FEATURES
source
1..590
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCLM0249A09"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UGCLM library"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gl14732114[9b]AP129072.1) a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance. "

ORIGIN
Query Match 80.8%; Score 20.2; DB 9; Length 590;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 AAAAAAAAACTAAGCTTGATCTTC 25
179 AAAAAAAAACTCAATCTTCATCTTC 203

RESULT 39
CL735418 618 bp DNA linear GSS 27-JUL-2004
LOCUS OR_BBa0068P10.f OR_BBa Oryza nivara genomic clone OR_BBa0068P10 5',
DEFINITION genomic survey sequence.
ACCESSION CL735418
VERSION CL735418.1 GI:50669774
KEYWORDS GSS.
SOURCE Oryza nivara
ORGANISM Oryza nivara
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 618)
Kim,H., Yu,Y., Scum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
OMAP Project
Unpublished (2004)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: twing@genome.arizona.edu

PCR primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Insert Length: 161 Std Error: 0.00

Plate: 0068 row: P column: 10

Seq primer: TAA TAC GAC TCA CTA TAG GG

Class: BAC ends.

FEATURES
Source Location/Qualifiers

1..618

/organism="Oryza nivara"

/mol_type="genomic DNA"

/db_xref="taxon:4536"

/clone="OR_B8a068P10"

/issue_type="Young Leaves"

/lab_host="DH10B-T1 phage resistant"

/clone_lib="OR_B8a"

/note="Vector: pAG1BAC1, Site_1: HindIII, Site_2: HindIII"

ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 10; Length 618;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAGCTTGATCTTC 25

Db 212 AAAAAAAAACTAATCTTCATCTTC 236

RESULT 40

CE071494

LOCUS CE071494 644 bp DNA linear GSS 24-SEP-2003

DEFINITION U1gr-ges-dog-1700033082385 Dog library Canis familiaris genomic,

genomic survey sequence.

CE071494

CE071494.1 GI:35133455

GSS.

Canis familiaris (dog)

Canis familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Laurasiatheria; Carnivora; Flesipedidae; Canidae;

Canis.

1 (bases 1 to 644)

Kirkness,E.F., Batina,V., Halpern,A.L., Levy,S., Remington,K.,

Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and

Venter,J.C.

The dog genome: survey sequencing and comparative analysis

Science 301 (5641), 1898-1903 (2003)

14512627

Contact: Kirkness EF

The Institute for Genomic Research

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

Rockville, MD 20850, USA

Tel: 301-838-0200

Fax: 301-838-0208

Email: ekirkness@tigr.org

Class: shotgun.

Location/Qualifiers

1..644

/organism="Canis familiaris"

/mol_type="genomic DNA"

/strain="Standard Poodle"

/db_xref="taxon:9615"

/clone_lib="Dog Library"

/note="Site 1: BactXI; Libraries were prepared from

peripheral blood"

ORIGIN

Query Match

Best Local Similarity 80.8%; Score 20.2; DB 9; Length 644;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAAAAAAAACTAAGCTTGATCTTC 25

Db 184 AAAAAAAAAAGAAAGCTTGATATTC 208

RESULT 41

AZ450483

LOCUS AZ450483 646 bp DNA linear GSS 04-OCT-2000

DEFINITION IM0249C15F Mouse 10kb plasmid UGCM library Mus musculus genomic

clone UGCM0249C15 F, genomic survey sequence.

AZ450483

AZ450483.1 GI:10605316

GSS.

Mus musculus (house mouse)

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,B., Pedersen,T.,

Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von

Niederhausen,A. and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0249 row: C column: 15

Seq primer: CGTTGTAAACGACGCCACAT

Class: plasmid ends

High quality sequence stop: 646.

Location/Qualifiers

1..646

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UGCM0249C15"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Ti-resistant, F-"

/clone_lib="mouse 10kb plasmid UGCM library"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (GI4732114|GB|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptor complementary to the insert adaptor and

purified. The sheared, adaptor mouse DNA was annealed to

adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

ORIGIN

Query Match

80.8%; Score 20.2; DB 9; Length 646;

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 663;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
 |||||
 Db 248 AAAAAAAAACTAATCTTCATCTTC 272

RESULT 45 702 bp mRNA linear EST 15-JUL-2005
 DR729413/LOCUS
 DEFINITION AGENCOURT 55147982 NICHD_Emb10 Xenopus laevis cDNA clone
 IMAGE:1982159 3', mRNA sequence.
 DR729413
 DR729413.1 GI:70903525

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Xenopus laevis (African clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Xenopus; Xenopus.
 1 (bases 1 to 702)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: csgarbs-remail.nih.gov

Tissue Procurement: Dr. Igor David
 cDNA Library Preparation: Express Genomics
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ULNL at:
 http://image.llnl.gov
 Plate: LHM1715 row: 0 column: 21
 High quality sequence start: 80
 High quality sequence stop: 698.

FEATURES

source

1..702
 Location/Qualifiers
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:7982159"
 /tissue_type="Embryonic stage 17/19"
 /lab_host="DH10B Tona"
 /clone_lib="NICHD_XGC_Emb10"
 /note="Organ: embryo; Vector: pExpress-1; Site: 1: EcoRV;
 Site: 2: NotI; cDNA was primed using oligo-dT primer:
 5'-GGAGTGTCTTAGATCGGAGCGGCCCT(T)25-3' and cloned into
 the EcoRV/NotI sites of pExpress-1. Size-selection 1.4kb
 resulted in an average insert size of 1.8kb. This is a
 normalized library (primary library is NICHD_XGC_Emb9) and
 was constructed by Express Genomics (Frederick, MD). Note:
 this is a Xenopus Gene Collection
 (http://xgc.nci.nih.gov/) library."

ORIGIN

Query Match 80.8%; Score 20.2; DB 8; Length 702;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
 |||||
 Db 355 AAAAAAAAAATACTTGATCTTC 331

RESULT 46

A2989040/c

LOCUS A2989040 703 bp DNA linear GSS 27-APR-2001
 DEFINITION 2M0272E16f Mouse 10kb plasmid U0GC2M library Mus musculus genomic
 clone U0GC2M0272E16 F, genomic survey sequence.

ACCESSION A2989040
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 703)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Dunn, P., Aoyagi, A., Barber, M., Baecorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Maimoud, M., Meenen, B., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von
 Niederhausern, A. and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)

Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
 Plate: 0272 row: E column: 16
 Seq primer: CGTGTAAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 703.

FEATURES

source

1..703
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U0GC2M0272E16"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv, Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD2 (g14732114|gb|AF129972.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 703;
 Best Local Similarity 88.0%; Pred. No. 3.2e+03;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
 |||||
 Db 76 AAAAAAAAAAGAGCTTGATCTTC 52


```

RESULT 47
AG349053      705 bp   DNA       linear   GSS 18-DEC-2004
LOCUS
DEFINITION   Mus musculus molossinus DNA, clone:MSMg01-145C23.TU, genomic survey
sequence.
ACCESSION    AG349053
VERSION      AG349053.1  GI:47922363
KEYWORDS
SOURCE       GSS.
ORGANISM     Mus musculus molossinus (Japanese wild mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiotoishi,T.
TITLE        Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL      Genome Res. 14 (12), 2439-2447 (2004)
PUBMED
REFERENCE
AUTHORS      Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyada, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector      : pBACe3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI.
FEATURES
source      Location/Qualifiers
1..705
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-145C23.TU"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
Query Match      80.8%; Score 20.2; DB 10; Length 705;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 92 AAAAAAACAATAATCTTGATCTTC 116
RESULT 48
CRJ13244/c      718 bp   DNA       linear   GSS 06-JUL-2004
LOCUS
DEFINITION   Forward strand read from insert in 3'HPRT insertion targeting and
chromosome engineering clone MHP159e06, genomic survey sequence.
ACCESSION    CRJ13244
VERSION      CRJ13244.1  GI:49878697
KEYWORDS
SOURCE       GSS; genome survey sequence; MICER.
Mus musculus (house mouse)

```

```

ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Meyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.
TITLE        Direct Submission
JOURNAL      Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES
source      Location/Qualifiers
1..718
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHP159e06"
/clone_lib="MHP"
ORIGIN
Query Match      80.8%; Score 20.2; DB 11; Length 718;
Best Local Similarity 88.0%; Pred. No. 3.2e+03;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 663 AATAACAATAAAGCTTATCTTC 639
RESULT 49
AGS71561      757 bp   DNA       linear   GSS 23-DEC-2004
LOCUS
DEFINITION   Mus musculus molossinus DNA, clone:MSMg01-494D17.T7, genomic survey
sequence.
ACCESSION    AGS71561
PUBMED
REFERENCE
AUTHORS      AGS71561
KEYWORDS
SOURCE       AGS71561.1  GI:4832360
GSS.
ORGANISM     Mus musculus molossinus (Japanese wild mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiotoishi,T.
TITLE        Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
JOURNAL      Genome Res. 14 (12), 2439-2447 (2004)
PUBMED
REFERENCE
AUTHORS      Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyada, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector      : pBACe3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI.
FEATURES
source      Location/Qualifiers

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2005, 02:43:33 ; Search time 180.2 Seconds
(without alignments)
68.002 Million cell updates/sec

Title: US-10-681-773-10

Perfect score: 25
Sequence: 1 aaaaaaaaaaagcttgatcttc 25

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4161359 seqs, 245077644 residues

Total number of hits satisfying chosen parameters: 8322718

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Published Applications NA_New:*

1: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.2	76.8	1756	6	US-10-750-185-3606
2	19.2	76.8	130733	7	US-11-121-086-19
3	19.2	76.8	171732	7	US-11-121-086-98
4	19.2	76.8	189539	7	US-11-121-086-16
5	18.6	74.4	201	6	US-10-995-561-62773
6	18.6	74.4	201	6	US-10-995-561-82633
7	18.6	74.4	1069	6	US-10-750-185-52544
8	18.6	74.4	2251	6	US-10-750-185-27339
9	18.6	74.4	2843	6	US-10-750-185-36379
10	18.6	74.4	27240	6	US-10-995-561-13499
11	18.6	74.4	52520	6	US-10-995-561-13360
12	18.6	74.4	85682	7	US-11-117-187-205
13	18.6	74.4	159146	7	US-11-121-086-45
14	18.6	74.4	182303	7	US-11-117-187-209
15	18.6	74.4	611587	7	US-10-995-561-47669
16	18.6	73.6	1433	6	US-10-750-185-40315
17	18.4	73.6	1599	6	US-10-750-185-28517
18	18.4	73.6	645179	6	US-10-995-561-12933
19	18.4	73.6	201	6	US-10-995-561-28493
20	18.2	72.8	905	6	US-10-750-185-54504
21	18.2	72.8	1597	6	US-10-750-185-42089
22	18.2	72.8	50353	6	US-10-995-561-13305
23	18.2	72.8	50353	6	US-10-995-561-13305

24	18.2	72.8	218821	7	US-11-121-086-11	Sequence 31, Appl
25	18.2	72.8	222094	6	US-10-995-561-13244	Sequence 13, Appl
26	17.6	70.4	201	6	US-10-995-561-76911	Sequence 76911, A
27	17.6	70.4	201	6	US-10-995-561-77084	Sequence 77084, A
28	17.6	70.4	201	6	US-10-995-561-77595	Sequence 77595, A
29	17.6	70.4	201	6	US-10-995-561-77596	Sequence 77596, A
30	17.6	70.4	447	6	US-10-793-626-823	Sequence 823, App
31	17.6	70.4	513	6	US-10-793-626-1953	Sequence 1953, Ap
32	17.6	70.4	600	6	US-10-750-185-2657	Sequence 2657, Ap
33	17.6	70.4	600	6	US-10-750-185-21354	Sequence 21354, A
34	17.6	70.4	1311	6	US-10-750-185-26879	Sequence 26879, A
35	17.6	70.4	1331	6	US-10-750-185-24894	Sequence 24894, A
36	17.6	70.4	1354	6	US-10-750-185-62604	Sequence 62604, A
37	17.6	70.4	1425	6	US-10-750-185-49713	Sequence 49713, A
38	17.6	70.4	1572	6	US-10-750-185-51637	Sequence 51637, A
39	17.6	70.4	1658	6	US-10-750-185-33931	Sequence 33931, A
40	17.6	70.4	1680	6	US-10-750-185-57535	Sequence 57535, A
41	17.6	70.4	1754	6	US-10-750-185-58714	Sequence 58714, A
42	17.6	70.4	1868	6	US-10-750-185-30064	Sequence 30064, A
43	17.6	70.4	1872	6	US-10-750-185-46372	Sequence 46372, A
44	17.6	70.4	2125	6	US-10-750-185-41395	Sequence 41395, A
45	17.6	70.4	2237	6	US-10-750-185-40001	Sequence 40001, A
46	17.6	70.4	2490	6	US-10-750-185-54317	Sequence 54317, A
47	17.6	70.4	2688	6	US-10-750-185-55769	Sequence 55769, A
48	17.6	70.4	2995	6	US-10-793-626-4272	Sequence 4272, Ap
49	17.6	70.4	3219	6	US-10-793-626-3405	Sequence 3405, Ap
50	17.6	70.4	3383	6	US-10-793-626-3363	Sequence 3363, Ap
51	17.6	70.4	3422	6	US-10-793-626-4400	Sequence 4400, Ap
52	17.6	70.4	4399	6	US-10-750-185-52104	Sequence 52104, A
53	17.6	70.4	40349	7	US-11-117-187-184	Sequence 184, App
54	17.6	70.4	173995	6	US-10-995-561-13277	Sequence 13277, A
55	17.6	70.4	191350	6	US-10-857-780-4	Sequence 4, Appl
56	17.6	70.4	191357	7	US-11-121-086-13	Sequence 13, Appl
57	17.6	70.4	260200	7	US-10-933-025-23	Sequence 23, Appl
58	17.6	70.4	340000	7	US-11-102-978-3	Sequence 3, Appl
59	17.6	70.4	387780	6	US-10-995-561-13475	Sequence 13475, A
60	17.6	70.4	394468	6	US-10-995-561-13473	Sequence 82062, A
61	17.2	68.8	201	6	US-10-995-561-82062	Sequence 82062, A
62	17.2	68.8	907	6	US-10-750-185-53785	Sequence 53785, A
63	17.2	68.8	1009	6	US-10-750-185-53776	Sequence 33776, A
64	17.2	68.8	1244	6	US-10-750-185-48909	Sequence 48909, A
65	17.2	68.8	1281	6	US-10-750-185-52883	Sequence 52883, A
66	17.2	68.8	1401	7	US-11-055-822-449	Sequence 449, App
67	17.2	68.8	1465	6	US-10-750-185-48726	Sequence 48726, A
68	17.2	68.8	1506	6	US-10-750-185-59676	Sequence 59676, A
69	17.2	68.8	1925	6	US-10-750-185-57026	Sequence 57026, A
70	17.2	68.8	43985	6	US-10-995-561-13337	Sequence 13337, A
71	17.2	68.8	150437	7	US-11-112-908-44	Sequence 44, Appl
72	17.2	68.8	150491	7	US-11-112-908-46	Sequence 46, Appl
73	17.2	68.8	182314	7	US-11-112-908-45	Sequence 45, Appl
74	17.2	68.8	186682	7	US-11-112-908-23	Sequence 23, Appl
75	17.2	68.0	201	6	US-10-995-561-12031	Sequence 32031, A
76	17.2	68.0	201	6	US-10-995-561-12524	Sequence 32524, A
77	17.2	68.0	201	6	US-10-995-561-16808	Sequence 36808, A
78	17.2	68.0	201	6	US-10-995-561-36932	Sequence 36932, A
79	17.2	68.0	201	6	US-10-995-561-33188	Sequence 33188, A
80	17.2	68.0	201	6	US-10-995-561-58454	Sequence 58454, A
81	17.2	68.0	201	6	US-10-995-561-80597	Sequence 80597, A
82	17.2	68.0	688	7	US-11-108-172-15	Sequence 15, Appl
83	17.2	68.0	808	6	US-10-750-185-47516	Sequence 47516, A
84	17.2	68.0	1098	6	US-10-750-185-46576	Sequence 46576, A
85	17.2	68.0	1140	6	US-10-750-185-60421	Sequence 60421, A
86	17.2	68.0	1157	6	US-10-750-185-60110	Sequence 60110, A
87	17.2	68.0	1303	6	US-10-750-185-58193	Sequence 58193, A
88	17.2	68.0	1406	6	US-10-750-185-33694	Sequence 33694, A
89	17.2	68.0	1648	6	US-10-750-185-42001	Sequence 42001, A
90	17.2	68.0	1654	6	US-10-750-185-36249	Sequence 36249, A
91	17.2	68.0	1728	6	US-10-750-185-49457	Sequence 49457, A
92	17.2	68.0	1728	6	US-10-750-185-50634	Sequence 50634, A
93	17.2	68.0	1852	6	US-10-750-185-55969	Sequence 55969, A
94	17.2	68.0	1933	6	US-10-750-185-77423	Sequence 37423, A
95	17.2	68.0	2095	7	US-11-102-240-35	Sequence 35, Appl
96	17.2	68.0	2248	6	US-10-750-185-25931	Sequence 25931, A

```

C 97 17 68.0 2713 6 US-10-750-185-32766 Sequence 32766, A
98 17 68.0 2791 6 US-10-750-185-55024 Sequence 55024, A
99 17 68.0 2967 6 US-10-750-185-33220 Sequence 33220, A
C 100 17 68.0 3001 7 US-11-145-703-183 Sequence 183, App
101 17 68.0 3086 6 US-10-750-185-63884 Sequence 63884, A
C 102 17 68.0 3233 6 US-10-750-185-25257 Sequence 25257, A
C 103 17 68.0 3298 6 US-10-750-185-63647 Sequence 63647, A
C 104 17 68.0 3820 6 US-10-750-185-64002 Sequence 64002, A
C 105 17 68.0 3935 6 US-10-750-185-36175 Sequence 36175, A
C 106 17 68.0 4135 6 US-10-523-912-3 Sequence 3, App11
C 107 17 68.0 8157 6 US-10-523-912-1 Sequence 1, App11
108 17 68.0 13682 6 US-10-995-561-13471 Sequence 13471, A
109 17 68.0 47622 6 US-10-995-561-13431 Sequence 13431, A
110 17 68.0 51917 6 US-10-995-561-13338 Sequence 13338, A
111 17 68.0 52192 6 US-10-995-561-13213 Sequence 13213, A
C 112 17 68.0 65835 6 US-10-995-561-13490 Sequence 13490, A
113 17 68.0 96539 6 US-10-995-561-13289 Sequence 13289, A
114 17 68.0 96583 6 US-11-117-187-203 Sequence 203, App
C 115 17 68.0 101001 6 US-10-995-561-13255 Sequence 13255, A
C 116 17 68.0 101786 6 US-11-117-187-199 Sequence 199, App
117 17 68.0 119036 6 US-10-995-561-13314 Sequence 13314, A
C 118 17 68.0 149382 6 US-10-995-561-13272 Sequence 13272, A
C 119 17 68.0 158410 7 US-11-121-086-46 Sequence 46, App1
120 17 68.0 162085 7 US-11-121-086-7 Sequence 7, App1
121 17 68.0 171936 6 US-10-933-025-24 Sequence 24, App1
C 122 17 68.0 180862 7 US-11-112-908-40 Sequence 40, App1
123 17 68.0 182314 7 US-11-112-908-45 Sequence 45, App1
C 124 17 68.0 196200 7 US-11-121-086-9 Sequence 9, App1
C 125 17 68.0 212716 7 US-11-121-086-95 Sequence 8, App1
C 126 17 68.0 246960 7 US-11-121-086-8 Sequence 11, App1
127 17 68.0 319608 7 US-11-145-703-1 Sequence 13204, A
C 128 17 68.0 321019 6 US-10-995-561-13204 Sequence 13259, A
C 129 17 68.0 387780 6 US-10-995-561-13259 Sequence 13274, A
C 130 17 68.0 415117 6 US-10-995-561-1274 Sequence 211, App
C 131 17 68.0 1082144 7 US-11-117-187-211 Sequence 637805, A
C 132 16.8 67.2 25 7 US-11-121-849-637805 Sequence 81867, A
C 133 16.8 67.2 201 6 US-10-995-561-81867 Sequence 82231, A
C 134 16.8 67.2 201 6 US-10-995-561-82231 Sequence 62, App1
135 16.8 67.2 420 6 US-10-927-641-62 Sequence 23680, A
136 16.8 67.2 821 6 US-10-750-185-25680 Sequence 47179, A
137 16.8 67.2 1599 6 US-10-750-185-47179 Sequence 43442, A
C 138 16.8 67.2 3175 6 US-10-750-185-43442 Sequence 127, App
139 16.8 67.2 3720 6 US-10-927-641-127 Sequence 3, App1
C 140 16.8 67.2 6450 7 US-11-091-668-3 Sequence 1353, A
141 16.8 67.2 103660 6 US-10-995-561-13253 Sequence 1, App1
142 16.8 67.2 126552 7 US-11-121-086-1 Sequence 6, App1
C 143 16.8 67.2 127343 7 US-11-121-086-6 Sequence 2, App1
C 144 16.8 67.2 191684 7 US-11-121-086-82 Sequence 82, App1
C 145 16.8 67.2 193084 7 US-11-121-086-82 Sequence 13286, A
C 146 16.8 67.2 1125001 6 US-10-995-561-31568 Sequence 31568, A
147 16.6 66.4 200 6 US-10-995-561-31568 Sequence 31640, A
148 16.6 66.4 201 6 US-10-995-561-31640 Sequence 31641, A
149 16.6 66.4 201 6 US-10-995-561-31641 Sequence 31642, A
150 16.6 66.4 201 6 US-10-995-561-35602 Sequence 35602, A
```

ALIGNMENTS

```

RESULT 1
US-10-750-185-36006
; Sequence 36006, Application US/10750185
; Publication No. US20050260603A1
;
GENERAL INFORMATION:
; APPLICANT: MMT GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFEID, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
```

```

; CURRENT APPLICATION NUMBER: US/10/750.185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36006
; LENGTH: 1756
; TYPE: DNA
; ORGANISM: Bovine 19866881236123
US-10-750-185-36006
```

```

Query Match 76.8%; Score 19.2; DB 6; Length 1756;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

Qy 2 AAAAAAAAACTAAGCTTGATCTTC 25
Db 972 AAAAAAAAACTCAAGCTTAATTTTC 995
```

```

RESULT 2
US-11-121-086-19
; Sequence 19, Application US/11121086
; Publication No. US20050266459A1
;
GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 19
; LENGTH: 130733
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-19
```

```

Query Match 76.8%; Score 19.2; DB 7; Length 130733;
Best Local Similarity 87.5%; Pred. No. 156+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

Qy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 49399 AAAAAAAAACTAATCTTGATCTT 49422
```

```

RESULT 3
US-11-121-086-98
; Sequence 98, Application US/11121086
; Publication No. US20050266459A1
;
GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 98
; LENGTH: 171732
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-98
```

```

Query Match 76.8%; Score 19.2; DB 7; Length 171732;
```

Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAAAGCTTGATCTT 24
Db 120362 AAAAAAAAAAAGCTTGATCTT 120385

RESULT 4
US-11-121-086-16
; Sequence 16, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 16
; LENGTH: 189539
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-16

Query Match 76.8%; Score 19.2; DB 7; Length 189539;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCTT 24
Db 188342 AAAAAAAAAAAGCTTGATCTT 188365

RESULT 5
US-10-995-561-62773
; Sequence 62773, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62773
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-62773

Query Match 74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 98;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCTT 25
Db 2 AAAAAAAAAAAGCTTGATCTT 26

RESULT 6
US-10-995-561-82633/c
; Sequence 82633, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82633
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-82633

Query Match 74.4%; Score 18.6; DB 6; Length 201;
Best Local Similarity 84.0%; Pred. No. 98;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCTT 25
Db 200 AAAAAAAAAAAGCTTGATCTT 176

RESULT 7
US-10-750-185-52544/c
; Sequence 52544, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52544
; LENGTH: 1069
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-52544

Query Match 74.4%; Score 18.6; DB 6; Length 1069;
Best Local Similarity 84.0%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGCTTGATCTT 25
Db 52 AAAAAAAAAAAGCTTGATCTT 28

RESULT 8
US-10-750-185-36679/c
; Sequence 36679, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185

CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 36679
LENGTH: 2251
TYPE: DNA
ORGANISM: Bovine 1986880665289
US-10-750-185-36679

Query Match 74.4%; Score 18.6; DB 6; Length 2251;
Best Local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 447 ACAAATAAATAAATCTTGTTCTTC 423

RESULT 9
US-10-750-185-27339
Sequence 27339, Application US/10750185
Publication No. US2005026063A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FATTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 27339
LENGTH: 2843
TYPE: DNA
ORGANISM: Bovine 1986880798157
US-10-750-185-27339

Query Match 74.4%; Score 18.6; DB 6; Length 2843;
Best Local Similarity 84.0%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 2674 AAAAAAGAGTCAAGCTTGATCTTC 2698

RESULT 10
US-10-995-561-13499/c
Sequence 13499, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CU001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13499
LENGTH: 27240
TYPE: DNA
ORGANISM: Homo sapiens

US-10-995-561-13499

Query Match 74.4%; Score 18.6; DB 6; Length 27240;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 24220 AAAAAAATAAAGTTTGACCTTC 24196

RESULT 11
US-10-995-561-13360
Sequence 13360, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
TITLE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CU001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13360
LENGTH: 52520
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) .. (52520)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13360

Query Match 74.4%; Score 18.6; DB 6; Length 52520;
Best Local Similarity 84.0%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 43683 AAAAAAATAAAGTTTGACCTTC 43707

RESULT 12
US-11-117-187-205
Sequence 205, Application US/11117187
Publication No. US20050266560A1
GENERAL INFORMATION:
APPLICANT: COPELAND, GREGORY
TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
FILE REFERENCE: ARCD:309US
CURRENT APPLICATION NUMBER: US/11/117,187
CURRENT FILING DATE: 2005-04-28
PRIOR APPLICATION NUMBER: US/09/531,120
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/125,219
PRIOR FILING DATE: 1999-03-18
NUMBER OF SEQ ID NOS: 212
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 205
LENGTH: 85682
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: modified base
LOCATION: (85362) .. (85481)
OTHER INFORMATION: N = A, C G, or T/U
US-11-117-187-205

Query Match 74.4%; Score 18.6; DB 7; Length 85682;
Best Local Similarity 84.0%; Pred. No. 2.4e+02;

```
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 50022 AAAAAAAAAAGAAAGCTTTTCTTC 50046

RESULT 13
US-11-121-086-49/c
; Sequence 49, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 49
; LENGTH: 159146
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-49

Query Match 74.4%; Score 18.6; DB 7; Length 159146;
Best Local Similarity 84.0%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 96285 AAAAAAAAACTAGACGATGATTTTC 96261

RESULT 14
US-11-121-086-45
; Sequence 45, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 45
; LENGTH: 182303
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-45

Query Match 74.4%; Score 18.6; DB 7; Length 182303;
Best Local Similarity 84.0%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 127102 AAAAAAAAAAAAGCTTTTCTTC 127126

RESULT 15
US-11-117-187-209
; Sequence 209, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE

; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD.30905
; CURRENT APPLICATION NUMBER: US/11/117,187
; PRIOR FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 209
; LENGTH: 611587
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-11-117-187-209

Query Match 74.4%; Score 18.6; DB 7; Length 611587;
Best Local Similarity 84.0%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTTC 25
Db 576248 AAAAAAAAAAGAAAGCTTTTCTTC 576272

RESULT 16
US-10-995-561-47669
; Sequence 47669, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47669
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-47669

Query Match 73.6%; Score 18.4; DB 6; Length 201;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTTGAT 21
Db 58 AAAAAAAAACTAAGCTTGAT 77

RESULT 17
US-10-750-185-40315
; Sequence 40315, Application US/10750185
; Publication No. US20050266060A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```

NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIN version 3.1
SEQ ID NO 40315
LENGTH: 1433
TYPE: DNA
ORGANISM: Bovine 19866881121145
US-10-750-185-40315

Query Match 73.6%; Score 18.4; DB 6; Length 1433;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGA 20
Db 184 AAAAAAAAACTAAAGCTTGA 203

RESULT 18
US-10-750-185-28517
Sequence 28517, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIN version 3.1
SEQ ID NO 28517
LENGTH: 1599
TYPE: DNA
ORGANISM: Bovine 19866881390037
US-10-750-185-28517

Query Match 73.6%; Score 18.4; DB 6; Length 1599;
Best Local Similarity 95.0%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 AAAAACTAAAGCTTGATCTTC 25
Db 950 AAAAACTAAATCTTGATCTTC 969

RESULT 19
US-10-995-561-13293
Sequence 13293, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: CL001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13293
LENGTH: 645179
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-13293

Query Match 73.6%; Score 18.4; DB 6; Length 645179;

Best Local Similarity 95.0%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAAGCTTGAT 21
Db 75327 AAAAAAAAACTAAAGCTTGAT 75346

RESULT 20
US-10-995-561-28493
Sequence 28493, Application US/10995561
Publication No. US20050272054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE REFERENCE: CL001559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28493
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-28493

Query Match 72.8%; Score 18.2; DB 6; Length 201;
Best Local Similarity 87.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCT 23
Db 95 AAAAAAAAACTAAAGCTTGATCT 117

RESULT 21
US-10-750-185-54504/c
Sequence 54504, Application US/10750185
Publication No. US20050260603A1
GENERAL INFORMATION:
APPLICANT: MMI GENOMICS, INC.
APPLICANT: DENISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIN version 3.1
SEQ ID NO 54504
LENGTH: 905
TYPE: DNA
ORGANISM: Bovine 19866881526848
US-10-750-185-54504

Query Match 72.8%; Score 18.2; DB 6; Length 905;
Best Local Similarity 87.0%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCT 23
Db 585 AAAAAAAAACTAAAGCTTGATCT 563

RESULT 22


```
US-10-750-185-42089/c
; Sequence 42089, Application US/10750185
; Publication No. US2005026063A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42089
; LENGTH: 1597
; TYPE: DNA
; ORGANISM: Bovine 19866880543152
US-10-750-185-42089

Query Match      72.8%; Score 18.2; DB 6; Length 1597;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 AAAAAAAAACTAAAGCTGATCT 24
Db 1373 AAAAAAAAACTAAAGCTGATCT 1351

RESULT 23
US-10-995-561-13305
; Sequence 561-13305, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13305
; LENGTH: 50353
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13305

Query Match      72.8%; Score 18.2; DB 6; Length 50353;
Best Local Similarity 87.0%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 AAAAAAAAACTAAAGCTGATCT 23
Db 49495 AAAAAAAAACTAAAGCTGATCT 49517

RESULT 24
US-11-121-086-31
; Sequence 31, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
```

```

; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 31
; LENGTH: 218821
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (106949)..(106949)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (110322)..(110324)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (115133)..(115133)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (133300)..(133300)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (139059)..(139158)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157740)..(157740)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157777)..(157777)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157900)..(157900)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157919)..(157919)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (157926)..(157926)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158094)..(158094)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158138)..(158138)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158193)..(158195)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158241)..(158242)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158259)..(158259)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (158278)..(158278)
; OTHER INFORMATION: a, c, g, t, unknown or other
```

```

FEATURE:
NAME/KEY: modified_base
LOCATION: (158295)..(158295)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158740)..(158839)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (158929)..(158929)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163528)..(163530)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163550)..(163550)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (163765)..(163765)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164000)..(164000)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164047)..(164047)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (164084)..(164084)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167233)..(167233)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167236)..(167236)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (167238)..(167238)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170247)..(170247)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170249)..(170250)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170252)..(170253)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170259)..(170259)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170262)..(170263)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (170266)..(170266)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:

```

```

NAME/KEY: modified_base
LOCATION: (174470)..(174470)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174472)..(174472)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (174474)..(174474)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179059)..(179060)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179064)..(179064)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197001)..(197001)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (197007)..(197007)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (200349)..(200349)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (200351)..(200351)
OTHER INFORMATION: a, c, g, t, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (212425)..(212426)
OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-31

```

```

Query Match 72.8%; Score 18.2; DB 7; Length 218821;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Oy 1 AAAAAAAAACTAAGCTTGATCT 23
Db 157187 AACCAAAATTAAGCTTGATTT 157209

```

```

RESULT 25
US-10-995-561-13244
Sequence 13244, Application US/10995561
Publication No. US2005027054A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
FILE OF INVENTION: DETECTION AND USES THEREOF
FILE REFERENCE: CLO01559
CURRENT APPLICATION NUMBER: US/10/995,561
CURRENT FILING DATE: 2004-11-24
NUMBER OF SEQ ID NOS: 85702
SOFTWARE: ParSeq for Windows Version 4.0
SEQ ID NO 13244
LENGTH: 222094
TYPE: DNA
ORGANISM: Homo sapiens
US-10-995-561-13244

```

```
Query Match          72.8%; Score 18.2; DB 6; Length 222094;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTGATCT 23
    ||||| ||||| ||||| |||||
Db 90939 AAAAAAAAAAGTAAAGCTGATCT 90961

RESULT 26
US-10-995-561-76911/c
; Sequence 76911, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76911
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-76911

Query Match          70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTGATCTTC 25
    ||||| ||||| ||||| |||||
Db 40 AAACAAACTAGATTGATCTTC 17

RESULT 27
US-10-995-561-77084/c
; Sequence 77084, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77084
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-77084

Query Match          70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTGATCTTC 25
    ||||| ||||| ||||| |||||
Db 175 AAACAAACTAGATTGATCTTC 152

RESULT 28
US-10-995-561-77595/c
; Sequence 77595, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77595
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-77595

Query Match          70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTGATCTTC 25
    ||||| ||||| ||||| |||||
Db 197 AAACAAACTAGATTGATCTTC 174

RESULT 29
US-10-995-561-77596/c
; Sequence 77596, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77596
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-77596

Query Match          70.4%; Score 17.6; DB 6; Length 201;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTGATCTTC 25
    ||||| ||||| ||||| |||||
Db 186 AAACAAACTAGATTGATCTTC 163

RESULT 30
US-10-793-626-823/c
; Sequence 823, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU048005
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 823
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
```

```
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-823

Query Match          70.4%; Score 17.6; DB 6; Length 447;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 2 AAAAAAAGCTTAAGCTTGATCTTC 25
    |||||
Db 372 AACATTAAGCTTAAGCTTGATCTTC 349

RESULT 31
US-10-793-626-1953/c
; Sequence 1953, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1953
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-1953

Query Match          70.4%; Score 17.6; DB 6; Length 513;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 2 AAAAAAAGCTTAAGCTTGATCTTC 25
    |||||
Db 438 AACATTAAGCTTAAGCTTGATCTTC 415

RESULT 32
US-10-750-185-2657/c
; Sequence 2657, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2657
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT07613
US-10-750-185-2657

Query Match          70.4%; Score 17.6; DB 6; Length 600;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Cy 2 AAAAAAAGCTTAAGCTTGATCTTC 25
    |||||
Db 156 AAAAAAAGCTTAAGCTTGATCTTC 133

RESULT 33
US-10-750-185-21354
; Sequence 21354, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21354
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT09140
US-10-750-185-21354

Query Match          70.4%; Score 17.6; DB 6; Length 600;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Cy 1 AAAAAAAGCTTAAGCTTGATCTTC 24
    |||||
Db 30 AAAAAAAGCTTAAGCTTGATCTTC 53

RESULT 34
US-10-750-185-26879
; Sequence 26879, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26879
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: Bovine 1986681611582
US-10-750-185-26879

Query Match          70.4%; Score 17.6; DB 6; Length 1311;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Db 336 AAAAAAAAAATAAGCTTCCTTT 359

RESULT 35

US-10-750-185-34894/c
; Sequence 34894, Application US/10750185
; Publication No. US20050260603A1

GENERAL INFORMATION:

APPLICANT: MMI GENOMICS, INC.

APPLICANT: DENISE, Sue K.

APPLICANT: KERR, Richard

APPLICANT: ROSENFELD, David

APPLICANT: HOLM, Tom

APPLICANT: BATES, Stephen

APPLICANT: FANTIN, Dennis

TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

FILE REFERENCE: MM1100-2

CURRENT APPLICATION NUMBER: US/10/750,185

PRIOR FILING DATE: 2003-12-31

PRIOR APPLICATION NUMBER: US 60/437,482

NUMBER OF SEQ ID NOS: 64922

SOFTWARE: PatentIn version 3.1

SEQ ID NO 34894

LENGTH: 1331

TYPE: DNA

ORGANISM: Bovine 19866881153190

US-10-750-185-34894

Query Match 70.4%; Score 17.6; DB 6; Length 1331;

Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 24

Db 894 AAAAAAAAAAGAGTGATCTT 871

RESULT 36

US-10-750-185-62604/c
; Sequence 62604, Application US/10750185
; Publication No. US20050260603A1

GENERAL INFORMATION:

APPLICANT: MMI GENOMICS, INC.

APPLICANT: DENISE, Sue K.

APPLICANT: KERR, Richard

APPLICANT: ROSENFELD, David

APPLICANT: HOLM, Tom

APPLICANT: BATES, Stephen

APPLICANT: FANTIN, Dennis

TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

FILE REFERENCE: MM1100-2

CURRENT APPLICATION NUMBER: US/10/750,185

PRIOR FILING DATE: 2003-12-31

PRIOR APPLICATION NUMBER: US 60/437,482

NUMBER OF SEQ ID NOS: 64922

SOFTWARE: PatentIn version 3.1

SEQ ID NO 62604

LENGTH: 1354

TYPE: DNA

ORGANISM: Bovine 19866880368620

US-10-750-185-62604

Query Match 70.4%; Score 17.6; DB 6; Length 1354;

Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 24

Db 199 AAAAAAAAAAGCTTAAGCTTGATTT 176

RESULT 37

US-10-750-185-49713/c
; Sequence 49713, Application US/10750185
; Publication No. US20050260603A1

GENERAL INFORMATION:

APPLICANT: MMI GENOMICS, INC.

APPLICANT: DENISE, Sue K.

APPLICANT: KERR, Richard

APPLICANT: ROSENFELD, David

APPLICANT: HOLM, Tom

APPLICANT: BATES, Stephen

APPLICANT: FANTIN, Dennis

TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

FILE REFERENCE: MM1100-2

CURRENT APPLICATION NUMBER: US/10/750,185

PRIOR FILING DATE: 2003-12-31

PRIOR APPLICATION NUMBER: US 60/437,482

PRIOR FILING DATE: 2002-12-31

NUMBER OF SEQ ID NOS: 64922

SOFTWARE: PatentIn version 3.1

SEQ ID NO 49713

LENGTH: 1425

TYPE: DNA

ORGANISM: Bovine 19866880556842

US-10-750-185-49713

Query Match 70.4%; Score 17.6; DB 6; Length 1425;

Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAAGCTTGATCTT 24

Db 501 AAGAAATCTAAAGCTTGATGTT 478

RESULT 38

US-10-750-185-51637/c
; Sequence 51637, Application US/10750185
; Publication No. US20050260603A1

GENERAL INFORMATION:

APPLICANT: MMI GENOMICS, INC.

APPLICANT: DENISE, Sue K.

APPLICANT: KERR, Richard

APPLICANT: ROSENFELD, David

APPLICANT: HOLM, Tom

APPLICANT: BATES, Stephen

APPLICANT: FANTIN, Dennis

TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

FILE REFERENCE: MM1100-2

CURRENT APPLICATION NUMBER: US/10/750,185

PRIOR FILING DATE: 2003-12-31

PRIOR APPLICATION NUMBER: US 60/437,482

PRIOR FILING DATE: 2002-12-31

NUMBER OF SEQ ID NOS: 64922

SOFTWARE: PatentIn version 3.1

SEQ ID NO 51637

LENGTH: 1572

TYPE: DNA

ORGANISM: Bovine 19866881195704

US-10-750-185-51637

Query Match 70.4%; Score 17.6; DB 6; Length 1572;

Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAAGCTTGATCTT 25

Db 780 AAAAAAAAAATAAGTTTCCTTTTC 757

RESULT 39

US-10-750-185-33931/c
; Sequence 33931, Application US/10750185

```
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33931
; LENGTH: 1658
; TYPE: DNA
; ORGANISM: Bovine 19866880421345
US-10-750-185-33931
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 1658;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 342 AAAAAAAAAAAAGCTTGCTCAT 319
```

```
RESULT 40
US-10-750-185-57535
; Sequence 57535, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57535
; LENGTH: 1680
; TYPE: DNA
; ORGANISM: Bovine 19866881410441
US-10-750-185-57535
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 1680;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 615 AAAAAAAAAATTAATCTTGATGTT 638
```

```
RESULT 41
US-10-750-185-58714/c
; Sequence 58714, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
```

```
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58714
; LENGTH: 1754
; TYPE: DNA
; ORGANISM: Bovine 19866880883020
US-10-750-185-58714
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 1754;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 945 AAAAAAAAAATTAATTAATCTTCTT 922
```

```
RESULT 42
US-10-750-185-30064
; Sequence 30064, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30064
; LENGTH: 1668
; TYPE: DNA
; ORGANISM: Bovine 19866881232408
US-10-750-185-30064
```

```
Query Match 70.4%; Score 17.6; DB 6; Length 1668;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 731 AAAAAAAAAAAAGCTTAATTTT 754
```

```
RESULT 43
US-10-750-185-46372/c
; Sequence 46372, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
```

```

; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46372
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Bovine 19866881061171
US-10-750-185-46372

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 6; Length 1872;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 186 AAAAAAAAACTAAGTTTITTTT 163

RESULT 44
US-10-750-185-41395
; Sequence 41395, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41395
; LENGTH: 2125
; TYPE: DNA
; ORGANISM: Bovine 19866880374636
US-10-750-185-41395

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 6; Length 2125;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 1887 AAAAAAAAACAACCTTCATCTT 1910

RESULT 45
US-10-750-185-40001
; Sequence 40001, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```

```

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40001
; LENGTH: 2237
; TYPE: DNA
; ORGANISM: Bovine 19866880291279
US-10-750-185-40001

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 6; Length 2237;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAAAAAAAACTAAGCTTGATCTT 24
Db 627 AAAAAAAAAAAAGCTTATCTT 650

RESULT 46
US-10-750-185-54317/c
; Sequence 54317, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54317
; LENGTH: 2490
; TYPE: DNA
; ORGANISM: Bovine 19866881057276
US-10-750-185-54317

Query Match
Best Local Similarity 83.3%; Score 17.6; DB 6; Length 2490;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAAAACTAAGCTTGATCTT 25
Db 481 AAAAAAAAACAACATGATCTGC 458

RESULT 47
US-10-750-185-55769/c
; Sequence 55769, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10750,185
```

CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 9.1
SEQ ID NO 55769
LENGTH: 2688
TYPE: DNA
ORGANISM: Bovine 19866880540328
US-10-750-185-55769

Query Match 70.4%; Score 17.6; DB 6; Length 2688;
Best Local Similarity 83.3%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAATAAGCTTGATCTTC 25
Db 428 AAAAAAATAAGCTTGATCTTC 405

RESULT 48
US-10-793-626-4272
Sequence 4272, Application US/10793626
Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/10/793,626
CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4272
LENGTH: 2995
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-4272

Query Match 70.4%; Score 17.6; DB 6; Length 2995;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAATAAGCTTGATCTTC 25
Db 1051 AAAAAAATAAGCTTGATCTTC 1074

RESULT 49
US-10-793-626-3405/C
Sequence 3405, Application US/10793626
Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/10/793,626
CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3405
LENGTH: 3219
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: nucleic acid sequence

US-10-793-626-3405

Query Match 70.4%; Score 17.6; DB 6; Length 3219;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAATAAGCTTGATCTTC 25
Db 62 AAAAAAATAAGCTTGATCTTC 39

RESULT 50
US-10-793-626-3363/C
Sequence 3363, Application US/10793626
Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/10/793,626
CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3363
LENGTH: 3383
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-3363

Query Match 70.4%; Score 17.6; DB 6; Length 3383;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAAAAAATAAGCTTGATCTTC 25
Db 2736 AAAAAAATAAGCTTGATCTTC 2713

Search completed: December 14, 2005, 11:41:02
Job time : 187.2 secs